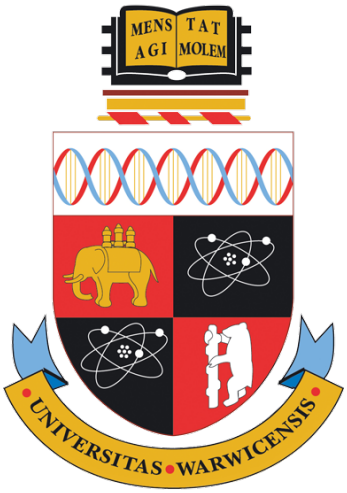

Point Estimation in Adaptive Confirmatory Clinical Trials with Time-to-Event Data and Treatment or Subgroup Selection

by

Josephine Naz Khan

Thesis

Submitted for the degree of

Doctor of Philosophy

Warwick Medical School

September 2017

Contents

List of Tables **v**

List of Figures **vi**

Acknowledgements **ix**

Declarations **x**

Abstract **xi**

Chapter 1 Introduction **1**

1.1 Outline of the thesis 3

Chapter 2 Clinical trials **6**

2.1 Phase I clinical trials 7

2.2 Phase II clinical trials 9

2.3 Phase III clinical trials 11

2.4 Adaptive seamless designs 12

2.4.1 Seamless phase II/III clinical trials 13

Chapter 3 Statistical background **16**

3.1 Probability density function 16

3.1.1 Multivariate distribution 17

3.2 Common distributions 18

3.3 Useful properties of the normal distribution 20

3.3.1 Standard normal 21

3.3.2 Truncated normal 22

3.3.3 Normal distribution as an example of an exponential family distribution 22

3.3.4 Multivariate normal distribution 23

3.3.5 Density of two univariate normals 24
3.4 Statistical properties of an estimator 27
3.4.1 Bias and mean squared error 27
3.4.2 Sufficiency and completeness 28
3.5 Minimum variance unbiased estimation 29

Chapter 4 Time-to-event data analysis 32
4.1 Introduction 32
4.2 Survival analysis 33
4.2.1 Censoring 34
4.2.2 Functions in survival data 35
4.2.3 Kaplan-Meier estimate of the survivor function 36
4.3 Hypothesis testing 38
4.3.1 Log-rank test 38
4.3.2 Wilcoxon test 40
4.3.3 Stratified log-rank test 41
4.4 Estimation of the log hazard ratio 41
4.4.1 Proportional hazards assumption 43
4.5 Modelling approaches 43
4.5.1 Cox proportional hazards model 44
4.5.2 Parametric models 45

Chapter 5 Unbiased estimation in seamless phase II/III clinical trials with normally distributed outcomes 48
5.1 Setting 49
5.2 Notation 49
5.3 Selection rule 50
5.4 Maximum likelihood estimators 50
5.5 Selection bias 51
5.5.1 Equal treatment means 53
5.5.2 Unequal treatment means 55
5.6 Unbiased estimation for normal outcomes 58
5.6.1 A UMVCUE accounting for treatment selection 59
5.7 Simulation study 64
5.8 Conclusion 67

Chapter 6 Correcting for treatment selection bias and correlation in seamless phase II/III clinical trials with time-to-event data 68

6.1 Setting . . . 69
6.2 Selection rule . . . 71
6.3 Simulation study assessing bias of the naive estimator . . . 72
6.3.1 Equal log HRs . . . 73
6.3.2 Unequal log HRs . . . 75
6.4 Independent increments structure . . . 77
6.5 Overview of available methods for analysing time-to-event data in seamless phase II/III clinical trials . . . 79
6.5.1 Joint distribution of log-rank test statistics . . . 80
6.5.2 Stage 2 increment statistics . . . 85
6.6 Derivation of the joint density of stagewise log HRs . . . 86
6.7 Bias correction with separate control arms . . . 88
6.8 Bias correction with a common control arm . . . 94
6.8.1 Bivariate asymptotic UMVCUE . . . 98
6.9 Simulation study . . . 99
6.9.1 Simulation set up . . . 99
6.9.2 Simulation results . . . 101
6.10 Conclusion . . . 112

Chapter 7 Subpopulation selection in adaptive survival trials 114
7.1 Setting . . . 114
7.1.1 Sources of bias . . . 117
7.2 Notation . . . 118
7.2.1 Estimating the log HRs . . . 118
7.3 Subpopulation selection rule . . . 120
7.4 Case where S is selected . . . 121
7.4.1 Naive estimator for θ_S . . . 121
7.4.2 An asymptotically UMVCUE for θ_S . . . 122
7.5 Case where F is selected . . . 126
7.5.1 Naive estimators . . . 127
7.5.2 Asymptotically unbiased estimators . . . 128
7.6 Simulation study . . . 129
7.6.1 Simulation set up . . . 129
7.6.2 Simulation results . . . 130
7.7 Conclusion . . . 131

Chapter 8 Summary 135
8.1 Further work . . . 140

Appendix A Proof of expectation of a truncated normal **143**

Appendix B Proof of sufficiency and completeness of $\tilde{\theta}^*$ **146**

Bibliography **155**

List of Tables

4.1 Data in each group at the j^{th} death time 39

5.1 Probability of selecting the truly most effective experimental treatment for normally distributed outcomes. Treatment arms are denoted by $k \in \{0, ..., K\}$, where $k = 0$ denotes the control arm. A range of treatment means $\mu_k \in \{0, ..., 0.6\}$ are considered for varying information fraction τ under 3 scenarios of equal effect sizes, one superior treatment, and unequal effect sizes. 56

6.1 Data at the j^{th} death time for up to K treatment groups 81
6.2 Summary of estimators for the selected log HR, θ_1 98
6.3 Mean bias and RMSE from 10,000 simulations of the naive estimator ($\hat{\theta}_1$), the stage 2 increment ($\tilde{\theta}_{21}$), and the UMVCUE ($\tilde{\theta}_{1b}$) for the case of two experimental treatment arms at selection times $\tau = \{0.3, 0.5, 0.8\}$. Log HRs θ_k $(k = 1, ..., 4)$ are assumed to be equal for each scenario. 109

7.1 Key notation 119
7.2 Mean bias and RMSE from 10,000 simulations of the naive estimator ($\hat{\theta}_S$) and the UMVCUE ($\tilde{\theta}_S$) for the case where S is selected. A range of log HRs θ_S are considered for subpopulation prevalence $p_S \in \{0.4, 0.5, 0.7\}$, selection times $\tau_S \in \{0.3, 0.5, 0.7\}$ and selection margin $c = 0$. 132
7.3 Mean bias and RMSE from 10,000 simulations of the naive estimator ($\hat{\theta}_F$) and the unbiased estimator ($\tilde{\theta}_F$) for the case where F is selected. A range of log HRs θ_F are considered for subpopulation prevalence $p_S \in \{0.4, 0.5, 0.7\}$, selection times $\tau_S \in \{0.3, 0.5, 0.7\}$ and selection margin $c = 0$. 133

List of Figures

2.1 Seamless phase II/III clinical trial schema 14

3.1 Normal density for μ of 0 (solid), -1 (dashed) and 1 (dotted) with $\sigma = 1$. 21

4.1 Kaplan-Meier estimate of simulated Weibull survival times for two treatment groups. Crosses indicate censored observations. 37
4.2 Weibull hazard function for $\gamma \in \{0.5, 1, 1.5, 2, 3\}$ with $\lambda = 1$. 46

5.1 Bias (top row) and RMSE (bottom row) depending on the number of experimental treatment arms $K \in \{1, ..., 6\}$ and the selection time $0 \leq \tau \leq 1$. Each column represents a different combination of treatment means, μ_k, $k \in \{0, ..., K\}$: (a) $\mu_k = 0$, (b) $\mu_1 = 0.3$ and $\mu_{k\neq 1} = 0$, and (c) $\mu_k = 0.1k$. 54
5.2 Contours of selection bias for $K = 3$ and $\tau = 1$ 57
5.3 Contours of RMSE for $K = 3$ and $\tau = 1$ 57
5.4 Bias (top row) and RMSE (bottom row) in units of $\sigma\sqrt{2/n}$ of $\hat{\theta}_s$ and $\tilde{\theta}_s$ for the scenario of 4 experimental treatment arms. Plots show the average of 500,000 simulations for each value of the selection time τ. Each column represents a different combination of treatment means μ_k $(k = 1, ..., 4)$: for equal means, $\mu_k = 0$; one superior mean, $\mu_1 = 0.3$ and $\mu_{k\neq 1} = 0$; and unequal means, $\mu_k = 0.1k$. 66

6.1 Example of survival data in a two-stage clinical trial. Black circles denote an event, and grey and white circles denote censored observations at the interim and final analysis, respectively. T_{interim} and T_{final} denote the calendar time of the interim and final analysis, respectively. 70

6.2 Bias (top row) and RMSE (bottom row) of the naive estimator for $K \in \{1, ..., 6\}$ treatment groups and selection time $\tau \in \{0, ..., 0.9\}$ from 10,000 simulations. Each column represents a different combination of effect sizes: (a) $\theta_k = 0\ \forall k \in \{1, ..., K\}$, (b) $\theta_k = -0.22\ \forall k \in \{1, ..., K\}$, and (c) $\theta_1 = -0.22$ and $\theta_k = 0\ \forall k \in \{2, ..., K\}$. . . . 76

6.3 Example of survival data in a two-stage clinical trial. Black circles denote an event, and grey and white circles denote censored observations at the interim and final analysis, respectively. T_{interim} and T_{final} denote the calendar time of the interim and final analysis, respectively. T_{follow} denotes the calendar time of maximum follow-up for patients recruited in stage 1. 78

6.4 Bias (top row) and RMSE (bottom row) of the stage 2 increment estimator $\tilde{\theta}_{21}$ for $K \in \{2, ..., 6\}$ treatment groups and selection times $\tau \in \{0.1, ..., 0.9\}$ from 10,000 simulations. Each column represents a different combination of effect sizes: (a) $\theta_k = 0\ \forall k \in \{1, ..., K\}$, (b) $\theta_k = -0.22\ \forall k \in \{1, ..., K\}$, and (c) $\theta_1 = -0.22$ and $\theta_k = 0\ \forall k \in \{2, ..., K\}$. 103

6.5 Bias (top row) and RMSE (bottom row) of the asymptotic UMVCUE for $K \in \{2, ..., 6\}$ treatment groups and selection times $\tau \in \{0.1, ..., 0.9\}$ from 10,000 simulations. Each column represents a different combination of effect sizes: (a) $\theta_k = 0\ \forall k \in \{1, ..., K\}$, (b) $\theta_k = -0.22\ \forall k \in \{1, ..., K\}$, and (c) $\theta_1 = -0.22$ and $\theta_k = 0\ \forall k \in \{2, ..., K\}$. . . . 104

6.6 Bias (top row) and RMSE (bottom row) of the naive estimator, stage 2 increment estimator, and asymptotic UMVCUE for a different number of treatment groups K and selection times $\tau \in \{0.1, ..., 0.9\}$, from 10,000 simulations. For each experimental treatment arm k, log HRs $\theta_k = 0\ \forall k \in \{1, ..., K\}$. 105

6.7 Bias (top row) and RMSE (bottom row) of the naive estimator, stage 2 increment estimator, and asymptotic UMVCUE for a different number of treatment groups K and selection times $\tau \in \{0.1, ..., 0.9\}$, from 10,000 simulations. For each experimental treatment arm k, log HRs $\theta_k = -0.22\ \forall k \in \{1, ..., K\}$. 106

6.8 Bias (top row) and RMSE (bottom row) of the asymptotic UMVCUE for a range of effect sizes from 10,000 simulations. A range of HRs (0.2,1) are assumed equal for K treatment arms and selection times $\tau \in \{0.1, ..., 0.9\}$. 110

6.9 Bias (top row) and RMSE (bottom row) of the naive estimator, stage 2 increment estimator, and asymptotic UMVCUE for unequal effect sizes for each treatment arm from 10,000 simulations. For each experimental treatment arm k, log HRs $\theta_k = -0.05k \ \forall k \in \{1, ..., K\}$. . 111

7.1 Population diagram. The full population is denoted by F, and the biomarker-positive and biomarker-negative subpopulations are denoted by S and S^c, respectively. 115

7.2 Two stage trial schema with subpopulation selection 116

Acknowledgements

I would like to acknowledge the financial support I received from the Medical Research Council and Novartis Pharma AG.

I am truly grateful to my academic supervisors, Peter Kimani and Nigel Stallard, for their advice and guidance through every stage of my PhD to completion of this thesis. Also, my sincere thanks to my industry partner supervisor, Ekkehard Glimm, for the stimulating research visits and opportunities at Novartis. I feel privileged to have worked with you all, and would like to thank you for sharing your time and knowledge with me.

To my parents, brothers, and friends far but close, thank you for your endless support and encouragement, not only during this PhD, but throughout my academic career to date; with special thanks to my friend Dani and former colleagues from the Birmingham CRCTU for inspiring me to undertake a PhD. Lastly, to my fellow PhDs at the Farmhouse and abroad, thank you for making this journey so memorable.

I dedicate this thesis to the memory of my grandfather, Karamat Hussain.

Declarations

I declare that this thesis is my own work, except where I have stated otherwise, and that it has not been submitted for a degree at any other university.

Abstract

Adaptive designs are increasingly adopted to make the process of drug development more efficient. In particular, seamless phase II/III clinical trials allow interim adaptations such as early stopping for futility or selection of the most promising treatment. Furthermore, targeted therapy trials include an interim analysis to select a subgroup with the largest observed treatment effect. However, despite their efficiency, data dependent adaptations lead to multiplicity and selection issues. This is because data are used for both treatment or subgroup selection as well as for the confirmatory analysis of treatment efficacy. Specifically, selection rules applied at the interim stage lead to overoptimistic and thus biased effect estimates. In this thesis, we investigate the bias that arises due to selection and develop unbiased estimators that correct for treatment or subgroup selection in two-stage confirmatory clinical trials with time-to-event data.

When analysis is based on time-to-event data, censoring at the interim analysis violates the assumption of independence between stage 1 and stage 2 data, which is a crucial assumption of existing methods for normally distributed data. The independent increments structure of stagewise log-rank test statistics has been beneficial for hypothesis testing in this setting, where group sequential methods have been utilised based on the log-rank test statistic for time-to-event data. We therefore incorporate the independent increments structure to derive unbiased estimators based on asymptotic normality of the log-rank test statistic. Additionally, when considering treatment selection, we address the issue of correlation between stage 1 estimates due to the common control arm for time-to-event outcomes.

We give the joint distribution of stagewise log hazard ratios and using the technique of Rao-Blackwellisation, we derive asymptotically uniformly minimum variance unbiased estimators conditional on selection rules for time-to-event outcomes. We examine the bias and mean squared error of conventional estimates and compare these, by simulation, to our unbiased and efficient estimators, which correct for treatment or subgroup selection and correlation due to both censoring and the common control arm. We show that, due to the asymptotic normality assumptions, our estimators are appropriate for large samples and small to moderate effect sizes.

Chapter 1

Introduction

The focus of this thesis is point estimation of treatment effects correcting for selection bias in two-stage confirmatory clinical trials with time-to-event data. Two forms of selection bias are explored. First, bias due to treatment selection in multi-arm seamless phase II/III clinical trials is assessed for both normally distributed outcomes and time-to-event outcomes, with the aim to develop new unbiased estimators for time-to-event outcomes. Second, bias due to subpopulation selection is investigated in the setting of two-stage targeted therapy trials, with the aim to develop new unbiased estimators.

In recent years, adaptive clinical trial designs have been adopted for use in drug development due to their ability to incorporate interim adaptations and test multiple hypotheses in a single trial [Hatfield et al., 2016]. In particular, seamless phase II/III clinical trials, which combine the goals of a phase II clinical trial with those of a phase III clinical trial, aim to accelerate the process of drug development by eliminating the time delay between conventionally separate phase II and III trials [Korn and Freidlin, 2017]. Much of the focus of statistical methodology for adaptive designs, specifically for seamless phase II/III clinical trials, has been on multiple

hypothesis testing [Bretz et al., 2006, Stallard and Todd, 2003]. Point estimation, on the other hand, specifically for two-stage trials with time-to-event data, seems to have attracted little attention. The overall aim of this thesis is therefore to explore the issues with time-to-event data analysis in this setting and develop unbiased estimators that correct for bias due to treatment or subpopulation selection.

In seamless phase II/III clinical trials, stage 1 involves investigation of multiple experimental treatments with the aim to select the most efficacious treatment at the interim analysis for further investigation in the second stage. This selection leads to biased estimation at the final analysis, since stage 1 data are used for both treatment selection and for the confirmatory analysis [Bauer and Posch, 2004]. In addition, correlation between stage 1 estimates arises as all experimental treatments are compared to a common control arm. In targeted therapy trials, selection with respect to a subgroup is made at the interim analysis based on the treatment effect observed in the subpopulation and the full population. As with treatment selection, this selection of the most promising population leads to biased estimation at the final analysis [Kimani et al., 2015]. Therefore, new unbiased estimators are warranted that utilise all trial data while considering both the selection rules and the correlation between estimates.

For normally distributed outcomes in seamless phase II/III clinical trials, unbiased estimators accounting for treatment selection exist for various trial scenarios. These include, early stopping for futility or efficacy [Kimani et al., 2013], and estimating treatment efficacy for not only the selected treatment but more than one experimental treatment [Bowden and Glimm, 2008]. These estimators utilise the technique of Rao-Blackwellisation, which involves conditioning on the selection rule and complete, sufficient statistics in order to derive a uniformly minimum variance conditionally unbiased estimator (UMVCUE). However, for time-to-event outcomes, such methods of efficient and unbiased estimation have not been investigated with

respect to selection bias. In targeted therapy trials, methods that address issues for hypothesis testing of time-to-event outcomes have been proposed [Glimm and Di Scala, 2015]. However, as with treatment selection, the problem of unbiased estimation accounting for subpopulation selection with time-to-event outcomes has been given less thought than that of hypothesis testing.

Time-to-event outcomes present additional challenges in estimation compared to normally distributed data due to the inherent nature of censoring. The main assumption of the estimators for normally distributed outcomes is that data from stage 1 and 2 are independent. With time-to-event outcomes, this assumption is violated, since at the interim analysis, the event of interest may not be observed for all patients recruited in stage 1. These patients are then followed-up further in stage 2, which therefore leads to correlated stage 1 and 2 data.

The estimators that will be developed in this thesis consider the issues that arise in the analysis of time-to-event data in two-stage clinical trials. Specifically, the issues that will be accounted for are the correlation between stagewise estimates and the bias, either due to treatment selection or subpopulation selection. Furthermore, derivation of point estimators will utilise the statistical technique of Rao-Blackwellisation in order to obtain asymptotically uniformly minimum variance conditionally unbiased estimators.

1.1 Outline of the thesis

An introduction to clinical trials is provided in Chapter 2 with an overview of trial designs for each phase of drug development. The setting of clinical trials for which the methods reported in Chapters 5 and 6 are based upon is then described.

Chapter 3 provides the statistical background for this thesis where standard statistical methods are given with a focus on point estimation. Further methods are

presented in Chapter 4 for the analysis of time-to-event data in the conventional setting of a single-stage two-arm clinical trial. Additionally, this chapter introduces the asymptotic normality assumptions for survival outcome parameters.

Chapter 5 presents existing methods of treatment effect estimation in seamless phase II/III clinical trials for normally distributed data. Bias that arises in maximum likelihood estimation due to treatment selection is examined analytically for various trial scenarios. An overview of current methods for unbiased estimation of normally distributed outcomes is provided before presenting a uniformly minimum variance conditionally unbiased estimator; this is then compared for efficiency to the biased maximum likelihood estimator via simulations.

Methods described in Chapter 5 are then extended in Chapter 6 to account for issues that arise in estimation of treatment effects in seamless phase II/III clinical trials with time-to-event outcomes. Selection bias from 'naive' estimation, which disregards selection and censoring, is assessed through simulations for various trial scenarios. Current methods that address multiplicity issues in time-to-event group sequential trials are presented with a view to using these methods to develop point estimators for time-to-event outcomes. Two new asymptotically UMVCUEs are then developed based on the log-rank test statistic. The first corrects for selection bias and correlation between stagewise estimates for the setting where separate control arms are assumed for each experimental treatment arm. The second new estimator derived replaces this assumption to correct for selection bias and correlation due to a common control arm. Results from a simulation study are then presented that compare the new unbiased estimator with a common control arm to the 'naive' estimator.

Chapter 7 addresses issues with treatment effect estimation in targeted therapy trials conditional on subpopulation selection. At the interim analysis, selection is made with respect to the biomarker-positive subpopulation or the full population. For

each case of selection, naive estimators are presented before deriving asymptotically unbiased estimators, conditional on selection. The estimators developed are then compared by simulation for a range of treatment effect sizes. All simulation results presented throughout this thesis have been programmed in the statistical language R [R Core Team, 2017].

This thesis concludes with chapter 8, which provides a summary of the work reported in this thesis and suggests areas for further research.

Chapter 2

Clinical trials

Clinical trials are prospective experimental studies conducted to assess the safety and efficacy of a healthcare intervention in human beings. An intervention may be a new medical treatment, a surgical procedure or a combination of existing treatments that are thought to interact. As this thesis is concerned with time-to-event outcomes, which are predominantly used in oncology trials, the interventions considered in this chapter are anti-tumour cytotoxic agents.

Traditional drug development consists of a sequence of independent phases of clinical trials, referred to as phase I−IV. In general, phase I and II clinical trials are considered the *learning phase* of drug development, whereas phase III and IV clinical trials make up the *confirming phase* [Sheiner, 1997]. Learning and confirmatory trials have different goals to progress drug development with individually tailored trial designs and methods of analysis. Phase IV clinical trials are post-licensing surveillance studies which are primarily interested in the long-term implementation of a drug. This thesis is interested in treatment effect estimation for pre-licensed drugs and therefore phase IV clinical trials are not discussed further in this thesis. The following sections give an overview of clinical trial designs for each phase

of development to provide the context of clinical trials that is the focus in this thesis.

2.1 Phase I clinical trials

The primary objective of phase I trials is to assess the safety profile of a drug whilst determining the biological optimal dose [Piantadosi, 2005]. These trials are known as 'first-in-man' studies as they are conventionally the first time a new treatment is introduced in humans for clinical assessment. They may also be the first time a combination of two or more existing treatments is explored for the joint safety profile and dosing regimen.

Phase I oncology trials are small sample studies generally conducted with patients who have tried and failed existing cancer therapies. For cytotoxic agents, it is believed that the dose-toxicity and dose-response relationship for outcomes of toxicity and efficacy are monotonically increasing. Therefore, the primary outcome measure of a phase I trial is the Maximum Tolerated Dose (MTD). This is considered the most effective tolerable dose and is estimated by observing the number of Dose Limiting Toxicities (DLTs) at different dose levels. Phase I studies follow a sequential dose-escalation design and can be either rule-based or model-based designs [George et al., 2016].

Traditionally in phase I oncology trials, rule-based designs such as the *3+3* and variations of this design were commonly used due to their simplicity and familiarity with clinical investigators. A prespecified set of dose levels is defined and allocation of 3 patients starts at the lowest dose level, with consecutive escalation of doses continued until the MTD is achieved. In this design, the MTD is defined as the highest dose level at which no more than 33% toxicities are observed [George et al., 2016]. If no DLTs occur at the current dose level, the dose is escalated and a new

cohort of 3 patients are recruited. If one or more toxicities occur at a given dose, then the dose is de-escalated for the subsequent cohort. This process is repeated until the MTD is found or all dose levels have been investigated. The escalation/de-escalation decision is only based on data from the current cohort of 3 patients at a given dose level. Hence, although commonly used, such rule-based designs are inefficient due to slow dose escalation and inefficient use of all trial data.

Alternatives to rule-based designs are model-based designs which explicitly fit parametric models for the dose-toxicity relationship. The most well-known model-based design is the *Continual Reassessment Method* (CRM) [O'Quigley et al., 1990]. This is a Bayesian adaptive design which is considerably more efficient both in terms of time and estimation of the MTD compared to all rule-based designs [Machin et al., 2004]. The most attractive feature of this design is that all accrued data are used in the escalation/de-escalation decision to determine the next dose level, as opposed to only the current cohort of three patients used in rule-based designs.

For the CRM, a target probability of toxicity is prespecified which may be higher or lower than the 33% in rule-based designs. The dose-toxicity curve is modelled after every patient is recruited and includes all accrued data to determine the next dose level from a set of dose levels. The next dose is estimated as the one whose probability of toxicity is closest to the target toxicity probability. Therefore, in comparison to rule based designs, this allows dose levels to be skipped in order to reach the MTD more quickly. The MTD is determined after a fixed number of patients have been treated. The model may take different forms, including: a power model, a hyperbolic tangent model or a logistic model. The main assumption of the model is that the dose-response relationship must be monotonically increasing, which is satisfied for cytotoxic agents. The CRM was initially adopted as a Bayesian model where prior probabilities of toxicities are elicited from clinical investigators to determine the starting dose level. For more details on model specification see

Conaway and Wages [2016].

Despite being attractive features of CRM, skipping of dose levels and updating the model after only one patient may not be desirable for safety and efficiency concerns [Green et al., 2002]. Modifications of the CRM have thus been proposed that address these concerns. To avoid time delays after each patient, the model may only be updated after a set number of patients have been observed for DLTs; for familiarity with rule-based designs, this may be set to 3 patients. Additionally, to avoid skipping to higher doses too quickly, the model may incorporate a restriction of skipping three or more dose levels, for example.

2.2 Phase II clinical trials

The second stage of drug development, defined as phase II, aims to determine an initial proof of activity of an experimental treatment. Phase II trials can be categorised into early phase II and late phase II trials depending on the stage of development. Early phase II oncology trials are usually single-arm non-randomised studies with small sample sizes [Machin et al., 2009], where the MTD found in phase I is brought forward into phase II to establish treatment activity whilst also monitoring toxicity. Late Phase II trials are randomised trials which aim to assess the activity of multiple experimental treatments simultaneously in order to select the most promising treatment to investigate further in phase III. As phase II trials are generally the first assessment of treatment efficacy, the primary outcome measure is usually a short objective measure and may be surrogate for the primary outcome in the subsequent phase III trial. For example, in phase II oncology trials, tumour response is typically considered a surrogate for overall survival time in a phase III trial.

Designs of phase II trials reflect the stage of development and depend on several factors including the primary outcome measure, the number of experimental treatments

and the number of stages. Early phase II designs are either single-arm single-stage designs [A'Hern, 2001, Fleming, 1982, Jung and Kim, 2004] or single-arm two-stage designs [Simon, 1989]. The two-stage designs incorporate an interim analysis to allow early stopping for inadequate activity. There are two variants of Simon's two-stage design; namely the minimax and optimal design. The minimax design aims to minimise the overall number of patients in the trial, whereas the optimal design aims to reduce the number of patients treated, should the treatment be found to be ineffective in stage 1. The optimal design is therefore favourable if the treatment is ineffective as it has a higher probability of stopping accrual in the second stage. On the other hand, if the treatment is found to be effective in stage 1, then the overall sample size would be larger than the minimax design. These designs follow a frequentist approach where critical values for desirable and undesirable levels of tumour response are defined for standard hypothesis testing. More details on these designs can found in Jung [2016].

Late phase II trial designs may be used to compare one or more experimental treatments or treatment doses, with the aim of selecting the most promising treatment/dose or dropping the least effective. Many randomised phase II designs have been proposed such as the 'pick a winner' design proposed by Simon et al. [1985]. Sargent and Goldberg [2001] proposed a more flexible design similar to Simon et al. that considers secondary endpoints when the observed treatment differences are small. Jung [2008] proposed an extension to Simon's [1989] two-stage single-arm design for randomised phase II trials with a control arm and one experimental arm. They provide a single-stage variation as well as the two-stage minimax and optimal designs. Rubinstein et al. [2005] provide a review for different phase II trial designs.

All the above designs assume a single primary outcome of activity which is usually the case in phase II trials. However, safety endpoints such as treatment tolerabil-

ity and toxicity are also monitored as secondary endpoints. The Bryant and Day [1995] design incorporates both safety and activity through a dual primary outcome measure of response and toxicity in a two-stage design. However, this design is only practically appropriate if response and toxicity assessments can be made in a similar time frame to determine the primary outcome of each patient, in order not to delay accrual in the second stage.

2.3 Phase III clinical trials

The final stage of drug development, before a treatment is considered for licensing, involves a large randomised, usually multi-centre, phase III clinical trial. Phase III clinical trials aim to provide a definitive assessment of treatment efficacy compared to a control treatment and are thus referred to as confirmatory clinical trials. The control treatment is generally the current standard of care for the disease in question. If no standard treatment exists then a placebo is generally used in order to provide a comparator arm for the experimental treatment.

Patients are randomly allocated to either the experimental treatment or control treatment, with allocation typically blinded to the patient and clinician. This is known as a double-blind trial and ensures allocation bias is minimised. Phase III clinical trials can take various designs depending on the aim of the treatment-control comparison and primary outcome measure. For example, sample size calculations vary for superiority aims, compared to equivalence or inferiority of treatment versus control [Green et al., 2002]. However, most commonly used sample size calculations are based on classical hypothesis testing with type II error rates minimised with a large sample size and confirmatory aims [Machin et al., 2009].

Trials with time-to-event outcomes are event-driven, which means statistical power and sample size calculations depend on the number of events as opposed to the

number of patients. Lakatos [2016] gives an overview of sample size calculations for survival trials in the context of oncology. Efficacy of treatments with survival outcomes are estimated in terms of a hazard ratio. This is a ratio of the death rates on the treatment and control arms. An overview of methods for time-to-event data analysis is provided in Chapter 4.

2.4 Adaptive seamless designs

As the combined goal of each phase of drug development is to determine whether an experimental treatment is safe and efficacious for use in patient care, in order to reduce the time delay between each study set up, it would be beneficial to combine these phases of clinical trials where appropriate. Adaptive seamless designs aim to eliminate the time delay between trials, whilst making more efficient use of patient data by incorporating objectives of multiple phases into a single, operationally and/or inferentially seamless trial. Operationally seamless trials make use of one trial protocol for the multiple stages, but data from each stage are analysed separately, while for inferentially seamless trials, analysis at the end of each stage includes all accrued data from previous stages.

The main feature of an adaptive design is the use of interim analyses. This allows design modifications to be made after each interim analysis. These modifications may include early stopping for futility or efficacy, sample size adjustments [Chuang-Stein et al., 2006] and selection of treatment arms [Posch et al., 2005]. Each interim analysis is conducted at the end of a stage in the trial, which therefore results in a multi-stage design. This multi-stage approach makes more efficient use of patient data, as the same data may be used to answer two or more clinical questions leading to reduced sample sizes compared to conducting separate trials.

However, incorporating such adaptations compromises the overall type I error rate

due to multiplicity issues. The main sources of multiplicity in confirmatory trials are repeated hypothesis testing at multiple interim analyses and investigating multiple experimental treatments for comparison with a common control group. Methods exist to control the type I error rate for these conditions known as 'multiple comparison procedures' and 'group sequential procedures' [Bretz et al., 2006, Jennison and Turnbull, 2000]. However, these methods require decision rules to be specified in advance. This is known as a 'prespecified adaptivity design' [Bretz et al., 2009] and is the setting considered in this thesis. For more details on multiplicity issues and multiple comparison procedures for multi-stage clinical trials see Xi et al. [2016].

Adaptive seamless designs that combine a late phase II trial and a phase III trial into one seamless trial with two stages, where stage 1 and 2 represent phase II and III components of a trial, are commonly referred to as *seamless phase II/III clinical trials* [Bretz et al., 2006]. Despite the attractive features of adaptive designs discussed above, the prespecified adaptivity requirement may eliminate the flexibility of modifying the phase III trial design, which may be desired based on results of the phase II trial. Furthermore, in addition to adjusting statistical tests to account for design modifications, adaptive trials require a greater time commitment both at the design and planning stage as well as from sponsoring organisations in contrast to an individual phase II trial that may not lead to a phase III trial [Korn and Freidlin, 2017]. Therefore, despite the many advantages of adaptive designs, the costs associated with such designs also need to be considered in practice.

2.4.1 Seamless phase II/III clinical trials

Seamless phase II/III clinical trials combine the aims of a late phase II and phase III trial into one operationally and inferentially seamless clinical trial [Korn et al., 2012]. The design of such a trial forms a two-stage multi-arm setting, where stage

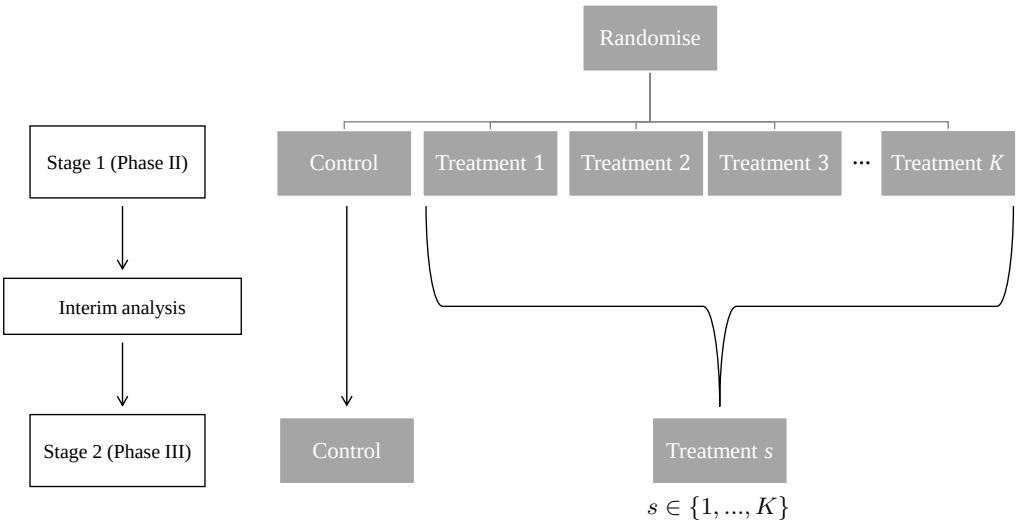

Figure 2.1: Seamless phase II/III clinical trial schema

1 addresses phase II objectives, investigating the activity of multiple treatments simultaneously, and stage 2 addresses the phase III objective regarding treatment efficacy. Figure 2.1 depicts this two-stage multi-arm setting.

In Chapters 5 and 6, like Stallard and Todd [2003], the phase II objective is defined as selection of the most promising treatment among several experimental treatments. In stage 1, patients are randomised to multiple experimental treatment arms plus a common control arm. At the end of stage 1, an interim analysis is performed where the 'best' performing experimental treatment is selected, according to a prespecified selection criterion, for further investigation in the stage 2 confirmatory phase III setting. Accrual continues in stage 2 where additional patients are randomised to the selected treatment and control arms. A final analysis is then conducted which utilises data accrued in both stages 1 and 2. This inferentially seamless two-stage approach is therefore said to be efficient in terms of data analysis as phase II data are not only used for the phase II objective but also for the phase III objective.

In addition to treatment selection, further design modifications may be incorporated at the interim analysis. These include early stopping for futility or efficacy [Stallard

and Todd, 2003] or adjusting the sample size for the subsequent stage [Bretz et al., 2006]. For example, if a treatment indicates inadequate efficacy at the interim analysis, the trial may stop early for futility. On the other hand, if the data show clear evidence of a beneficial effect, further patients may not be required. Hence, combining both phase II and phase III trial aims into one seamless phase II/III trial makes drug development efficient, as it is not only time- and cost-effective, but also answers multiple clinical questions with potentially smaller sample sizes.

Therefore, seamless phase II/III trials are attractive for use in multi-arm clinical trials. However, due to the interim analysis, challenges arise in statistical inference. Estimation following treatment selection is biased due to multiple use of stage 1 data. Existing methods for unbiased estimation for normally distributed outcomes that correct for such selection bias are presented in Chapter 5. However, these methods cannot be used in the analysis of time-to-event data as such data are not normally distributed and include censored event times. Therefore, Chapters 6 and 7 develop methods for treatment effect estimation that correct for selection bias which are appropriate for time-to-event outcomes.

Chapter 3

Statistical background

There are three main areas of statistical inference known as *point estimation, confidence interval estimation* and *hypothesis testing*. This thesis is concerned with point estimation in confirmatory clinical trials, and therefore this chapter introduces statistical inference tools that are used throughout this thesis, particularly focussing on point estimation. The methods described in this chapter can be found in theoretical statistics textbooks such as Lindgren [1993], Roussas [2007] and Young and Smith [2005]. This chapter begins by introducing probability distributions considered in this thesis, followed by properties of the normal distribution relevant to later chapters, and ends with statistical properties of point estimators.

3.1 Probability density function

Suppose X is a random variable which has probability distribution that depends on an unknown parameter θ. Let x denote the observed value of X, then the probability distribution of X is specified by a *probability density function* (pdf) if X is continuous or a *probability mass function* (pmf) if X is discrete. This is denoted

by $f_\theta(x)$, where the index indicates the dependence on θ. The probability that X is less than or equal to x is described by the *cumulative distribution function* (cdf), $F_\theta(x)$.

In parametric inference, $f_\theta(x)$ is of a known analytic form, whereas non-parametric inference does not assume the data follow a known statistical distribution. Chapter 4 alludes to semi-parametric inference but otherwise we are only concerned with parametric inference in this thesis.

3.1.1 Multivariate distribution

A function $f(x_1, ..., x_n)$ from $\mathbb{R}^n$ to $\mathbb{R}$ is called a *joint pdf* of the continuous random vector $(X_1, ..., X_n)$ if, for any subset $\mathcal{A} \subset \mathbb{R}^n$,

$$P((X_1, ..., X_n) \subset \mathcal{A}) = \int \dots \int_{\mathcal{A}} f(x_1, ..., x_n) \mathrm{d}x_1 \dots \mathrm{d}x_n.$$

The *joint cdf* of $X_1, ..., X_n$ is thus

$$F(x_1, ..., x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f(u_1, ..., u_n) \mathrm{d}u_1 \dots \mathrm{d}u_n \quad \forall (x_1, ..., x_n) \in \mathbb{R}^n.$$

Transformation of random variables

Transformation of random variables is an important technique that allows the distribution of a desired random vector to be found from a known distribution of a related random vector. This tool will be used in later chapters and is thus described in general below. Assuming the relation is a *one-to-one* multivariate transformation, the density of the transformed joint distribution is found as follows.

Let $X_1, ..., X_n$ be random variables with joint pdf $f(x_1, ..., x_n)$. Let $g(y_1, ..., y_n)$ be the joint pdf of $Y_1, ...Y_n$, where $Y_i = h_i(X_1, ..., X_n)$ for $i = 1, ..., n$ and h_i are

invertible functions. Suppose the equations y_i can be solved for $x_i = h_i^{-1}(y_1, ..., y_n)$ and the inverse functions h_i^{-1} are differentiable. Then the joint pdf of $Y_1, ..., Y_n$ is

$$f(y_1, ..., y_n) = f(h_1^{-1}(y_1, ..., y_n), ..., h_n^{-1}(y_1, ..., y_n))|J|, \tag{3.1}$$

where $|J|$ is the absolute value of the determinant of the Jacobian of the transformation

$$J = \begin{vmatrix} \frac{\delta h_1^{-1}}{\delta y_1} & \cdots & \frac{\delta h_1^{-1}}{\delta y_n} \\ \vdots & & \vdots \\ \frac{\delta h_n^{-1}}{\delta y_1} & \cdots & \frac{\delta h_n^{-1}}{\delta y_n} \end{vmatrix}.$$

3.2 Common distributions

This section gives general properties of statistical distributions considered in this thesis. Probability density functions, $f(x)$, are given for continuous distributions while probability mass functions, $P(X = x)$, are given for discrete distributions, with their respective expected value, $\mathrm{E}(X)$, and variance, $\mathrm{var}(X)$.

The hypergeometric distribution: $Hyp(N, r, n)$

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, ..., n; N, r, n \in \mathbb{N}_{>0}. \tag{3.2}$$

Let $p = \frac{r}{N}$, then

$$\begin{aligned} \mathrm{E}(X) &= np; \\ \mathrm{var}(X) &= \frac{N-n}{N-1} np(1-p). \end{aligned}$$

If $N \to \infty$, then $\mathrm{var}(X)$ converges to $np(1-p)$.

The normal distribution: $N(\mu, \sigma^2)$

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \quad x, \mu \in \mathbb{R}, \sigma^2 > 0. \\ \mathrm{E}(X) &= \mu; \\ \mathrm{var}(X) &= \sigma^2. \end{aligned} \tag{3.3}$$

The normal distribution, also called the Gaussian distribution, has many useful properties as described in the next section.

The chi-squared distribution: χ^2_μ

$$\begin{aligned} f(x) &= \frac{1}{\Gamma(\mu/2)2^{\mu/2}} x^{(\mu/2)-1} e^{-x/2}, \quad x \geq 0, \mu \in \mathbb{N}_{>0} \\ \mathrm{E}(X) &= \mu; \\ \mathrm{var}(X) &= 2\mu. \end{aligned} \tag{3.4}$$

If $X_1, ..., X_n$ are independent random variables with $X_i \sim N(\mu, \sigma^2)$, then $\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$.

The uniform distribution: $U(a, b)$

$$\begin{aligned} f(x) &= \begin{cases} \dfrac{1}{b-a}, & \text{if } x \in [a, b]; a, b \in \mathbb{R} \\ 0, & \text{otherwise.} \end{cases} \\ \mathrm{E}(X) &= \frac{b-a}{2}; \\ \mathrm{var}(X) &= \frac{(b-a)^2}{12}. \end{aligned} \tag{3.5}$$

The Weibull distribution: $Weibull(\mu, \sigma)$

$$\begin{aligned} f(x) &= \mu\sigma x^{\sigma-1} e^{-\mu x^{\sigma}}, \quad x, \mu, \sigma > 0. \\ \mathrm{E}(X) &= \mu^{-1/\sigma}\Gamma\left(1+\frac{1}{\sigma}\right); \\ \mathrm{var}(X) &= \mu^{-2/\sigma}\Gamma\left(1+\frac{2}{\sigma}\right). \end{aligned} \tag{3.6}$$

This is the common parametrisation used in medical research. If $\mu = 1$ then this reduces to the exponential distribution.

Exponential families

Some of the distribution functions given above are special cases of exponential families. Suppose a random vector of n independent observations, $\boldsymbol{X} = (X_1, ..., X_n)$, depends on the parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$. Then a family of pdfs or pmfs of $\boldsymbol{X}$ is called a $p-$parameter exponential family if its joint pdf or pmf can written in the form

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \prod_{i=1}^{n} h(x_i) c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{p} q_j(\boldsymbol{\theta}) t_j(x_i)\right). \tag{3.7}$$

An example of this is given for the normal distribution in Section 3.3.3.

3.3 Useful properties of the normal distribution

The normal distribution has a focal role in probability and statistics for three main reasons [Casella and Berger, 2002]. Firstly, the distribution is analytically tractable which means exact analytical solutions can be found easily. Secondly, the distribution has a symmetric bell-shaped density, which makes it a highly applicable choice

for population models. The third, and perhaps most important reason, is its use as an approximation to other distributions for large samples via the Central Limit Theorem [Lindgren, 1993].

3.3.1 Standard normal

If a random variable X has a normal distribution with mean μ and variance σ^2 then it is common to write $X \sim N(\mu, \sigma^2)$, with density $f(x)$ given in equation (3.3). If $X \sim N(\mu, \sigma^2)$, then the z-score

$$Z = \frac{X - E(X)}{\sqrt{\text{var}(X)}}$$

is said to be standard normal. This is commonly written as $Z \sim N(0, 1)$. The pdf and cdf of a standard normal variable are given by $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ and $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} \mathrm{d}u$, respectively. Thus, the density and cumulative distribution function of any normally distributed random variable X with mean μ and variance σ^2 can be expressed in terms of the standard normal distribution by $f(x) = \frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})$ and $F(x) = \Phi(\frac{x-\mu}{\sigma})$. Figure 3.1 shows the bell shape of the normal density function with symmetry around its mean. This implies $\phi(x) = \phi(-x)$ and $\Phi(-x) = 1 - \Phi(x)$.

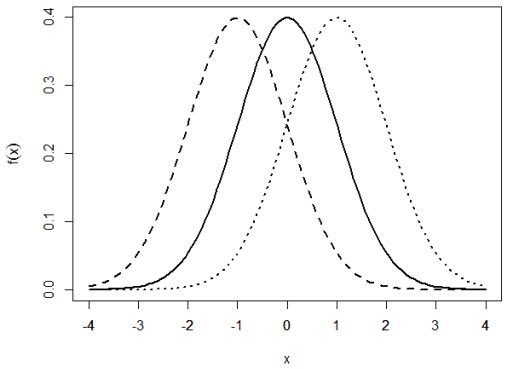

Figure 3.1: Normal density for μ of 0 (solid), -1 (dashed) and 1 (dotted) with $\sigma = 1$.

3.3.2 Truncated normal

The range of a normal random variable may be restricted above or below which leads to a truncated normal distribution. The expectation of a truncated normal pdf is of interest in later chapters and is thus given here. The expectation of $X \sim N(\mu, \sigma^2)$ is found by $\int_{-\infty}^{\infty} \frac{x}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \mathrm{d}x$, which implies

$$\begin{aligned} \int_{-\infty}^{b} \frac{x}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \mathrm{d}x + \int_{b}^{\infty} \frac{x}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \mathrm{d}x &= \mu \\ \implies \int_{b}^{\infty} \frac{x}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \mathrm{d}x &= \mu - \int_{-\infty}^{b} \frac{x}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \mathrm{d}x. \end{aligned} \tag{3.8}$$

The mean of an upper tail truncated normal random variable is given by

$$\int_{-\infty}^{b} \frac{x}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \mathrm{d}x = -\sigma\phi\left(\frac{b-\mu}{\sigma}\right) + \mu\Phi\left(\frac{b-\mu}{\sigma}\right), \tag{3.9}$$

as shown in Appendix A.

3.3.3 Normal distribution as an example of an exponential family distribution

It can be shown that the density function of the normal distribution with unknown mean and variance belongs to a two parameter exponential family. Recall the density function given in equation (3.3). Following equation (3.7), let $h(x) = 1$, $c(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\mu^2}{2\sigma^2})$, $q_1(\mu, \sigma^2) = -\frac{1}{2\sigma^2}$, $t_1(x) = x^2$, $q_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2}$ and $t_2(x) = x$, then the normal density function can be expressed as

$$f(x) \quad = \quad \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left[-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x\right]. \tag{3.10}$$

Hence, from this parameterisation and Definition 3.4.8, the complete, sufficient statistic for (μ, σ^2) is

$$T(\boldsymbol{X}) \quad = \quad \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2 \right).$$

Let $\bar{X}$ and S^2 denote the normal sample mean and standard deviation. Then the one-to-one transformation of $T(\boldsymbol{X})$ to $T_1(\boldsymbol{X}) = (\bar{X}, S^2)$, implies $T_1(\boldsymbol{X})$ is complete and sufficient for (μ, σ^2), and also minimal sufficient as it is two dimensional. These concepts are formally defined in Section 3.4.2.

3.3.4 Multivariate normal distribution

A $p-$dimensional random vector $\boldsymbol{X} = (X_1, ..., X_p)'$ has a $p-$variate normal distribution if the joint pdf of $\boldsymbol{X}$ is

$$f_{\boldsymbol{X}}(x) = \frac{1}{(2\pi)^{n/2}\sqrt{|\boldsymbol{V}|}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\}$$

where $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)'$ and $\boldsymbol{V}$ is the positive definite variance-covariance matrix.

If $X_1, ..., X_p$ are *independent* univariate normal $N(\mu_i, \sigma_i^2)$, then $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{V})$ where $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)'$ and $\boldsymbol{V} = diag(\sigma_1^2, ..., \sigma_p^2)$ is a diagonal, positive definite covariance matrix, where $\text{cov}(X_i, X_j) = 0$. This can be written as

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \sim N_p \left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p^2 \end{pmatrix} \right).$$

Conditional distribution

Now suppose $\boldsymbol{X}$ is partitioned into $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, where $\boldsymbol{X}_1 = (X_1, ..., X_q)$ is a $q-$dimensional multivariate normal with mean vector $\boldsymbol{\mu}_1 = (\mu_1, ..., \mu_q)$ and covariance matrix $\boldsymbol{V}_1$, and $\boldsymbol{X}_2 = (X_{q+1}, ..., X_p)$ is a $(p-q)-$dimensional multivariate normal with mean vector $\boldsymbol{\mu}_2 = (\mu_{q+1}, ..., \mu_p)$ and covariance matrix $\boldsymbol{V}_2$. Then we can write the joint multivariate normal distribution of $(\boldsymbol{X}_1, \boldsymbol{X}_2)'$ as

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_p \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_2 \end{pmatrix} \right)$$

where $\boldsymbol{V}_{12}$ and $\boldsymbol{V}_{21}$ are $\text{cov}(\boldsymbol{X}_1, \boldsymbol{X}_2)$.

By standard theory of conditional probabilities, the conditional distribution of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2 = \boldsymbol{x}_2$, is defined by

$$f(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f(\boldsymbol{x}_1)} \quad \forall \boldsymbol{x}_1 : f(\boldsymbol{x}_1) > 0 \text{ and } \forall \boldsymbol{x}_2 \in \mathbb{R}.$$

Hence, it can be shown that the conditional distribution of $\boldsymbol{X}_1|\boldsymbol{X}_2$ is $q-$dimensional multivariate normal with conditional mean and variance given by $\boldsymbol{\mu}_1 + \boldsymbol{V}_{12}\boldsymbol{V}_2^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$ and $\boldsymbol{V}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_2^{-1}\boldsymbol{V}_{21}$, respectively. This can be written as

$$\boldsymbol{X}_1|\boldsymbol{X}_2 \sim N_q(\boldsymbol{\mu}_1 + \boldsymbol{V}_{12}\boldsymbol{V}_2^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{V}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_2^{-1}\boldsymbol{V}_{21}).$$

3.3.5 Density of two univariate normals

The joint density of two univariate normals with equal means is required for the derivation of point estimators in later chapters. This section gives detailed steps for re-expressing this density to allow for more coherent reading in later chapters.

Let $X \sim N(\mu, \sigma_1^2)$ and $Y \sim N(\mu, \sigma_2^2)$, we want to show

$$\frac{1}{\sigma_1}\phi\left(\frac{x-\mu}{\sigma_1}\right)\frac{1}{\sigma_2}\phi\left(\frac{y-\mu}{\sigma_2}\right) = \phi\left(\frac{\alpha_1\mu - (\frac{\sigma_2}{\sigma_1}x + \frac{\sigma_1}{\sigma_2}y)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)\frac{1}{\sigma_1^2}\phi\left(\frac{x - (\frac{\sigma_2}{\sigma_1}x + \frac{\sigma_1}{\sigma_2}y)/\alpha_1}{\frac{\sigma_1^2}{\sigma_2^2}\alpha_2}\right) \tag{3.11}$$

where $\alpha_1 = \frac{\sigma_1}{\sigma_2} + \frac{\sigma_2}{\sigma_1}$ and $\alpha_2 = \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}$.

Writing out the density functions gives

$$\begin{aligned}
& \phi\left(\frac{x-\mu}{\sigma_1}\right)\phi\left(\frac{y-\mu}{\sigma_2}\right) \\
= \;& \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma_1^2}(x^2 - 2\mu x + \mu^2)\right\}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma_2^2}(y^2 - 2\mu y + \mu^2)\right\} \\
= \;& \frac{1}{2\pi}\exp\left\{-\frac{1}{2\sigma_1^2\sigma_2^2}\left[\sigma_2^2(x^2 - 2\mu x + \mu^2) + \sigma_1^2(y^2 - 2\mu y + \mu^2)\right]\right\}. \qquad (3.12)
\end{aligned}$$

Now expanding and factorising the terms inside the square brackets gives

$$\begin{aligned}
& \sigma_2^2x^2 - 2\mu\sigma^2x + \mu^2\sigma_2^2 + \sigma_1^2y^2 - 2\mu\sigma_1^2y + \mu^2\sigma_1^2 \\
= \;& \mu^2(\sigma_1^2+\sigma_2^2) - 2\mu(\sigma_2^2x + \sigma_1^2y) + \sigma_2^2x^2 + \sigma_1^2y^2 \\
= \;& (\sigma_1^2+\sigma_2^2)\left\{\mu^2 - \frac{2\mu(\sigma_2^2x + \sigma_1^2y)}{\sigma_1^2+\sigma_2^2} + \frac{\sigma_2^2x^2 + \sigma_1^2y^2}{\sigma_1^2+\sigma_2^2}\right\} \\
= \;& (\sigma_1^2+\sigma_2^2)\left\{\left(\mu - \frac{\sigma_2^2x + \sigma_1^2y}{\sigma_1^2+\sigma_2^2}\right)^2 - \left(\frac{\sigma_2^2x + \sigma_1^2y}{\sigma_1^2+\sigma_2^2}\right)^2 + \frac{\sigma_2^2x^2 + \sigma_1^2y^2}{\sigma_1^2+\sigma_2^2}\right\} \\
= \;& (\sigma_1^2+\sigma_2^2)\left\{\left(\mu - \frac{\sigma_2^2x + \sigma_1^2y}{\sigma_1^2+\sigma_2^2}\right)^2 - \frac{(\sigma_2^2x + \sigma_1^2y)^2}{(\sigma_1^2+\sigma_2^2)^2} + \frac{\sigma_2^2x^2 + \sigma_1^2y^2}{\sigma_1^2+\sigma_2^2}\cdot\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right\} \\
= \;& (\sigma_1^2+\sigma_2^2)\left\{\left(\mu - \frac{\sigma_2^2x + \sigma_1^2y}{\sigma_1^2+\sigma_2^2}\right)^2 + \frac{\sigma_1^2\sigma_2^2x^2 - 2\sigma_1^2\sigma_2^2xy + \sigma_1^2\sigma_2^2y^2}{(\sigma_1^2+\sigma_2^2)^2}\right\} \\
= \;& (\sigma_1^2+\sigma_2^2)\left\{\left(\mu - \frac{\sigma_2^2x + \sigma_1^2y}{\sigma_1^2+\sigma_2^2}\right)^2 + \frac{\sigma_1^2\sigma_2^2}{(\sigma_1^2+\sigma_2^2)^2}(x-y)^2\right\}.
\end{aligned}$$

Substituting this back into (3.12) leads to

$$\begin{aligned}
& \frac{1}{2\pi}\exp\left\{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left\{\left(\mu-\frac{\sigma_2^2x+\sigma_1^2y}{\sigma_1^2+\sigma_2^2}\right)^2+\frac{\sigma_1^2\sigma_2^2}{(\sigma_1^2+\sigma_2^2)^2}(x-y)^2\right\}\right\} \\
= & \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(\mu-\frac{\sigma_2^2x+\sigma_1^2y}{\sigma_1^2+\sigma_2^2}\right)^2\right\}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x-y)^2}{2(\sigma_1^2+\sigma_2^2)}\right\} \\
= & \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\frac{\sigma_1^2+\sigma_2^2}{\sigma_1\sigma_2}\mu-\frac{\sigma_2^2x+\sigma_1^2y}{\sigma_1\sigma_2}}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)^2\right\}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x-y)^2}{2(\sigma_1^2+\sigma_2^2)}\right\} \\
= & \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\alpha_1\mu-(\frac{\sigma_2}{\sigma_1}x+\frac{\sigma_1}{\sigma_2}y)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)^2\right\}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x-y)^2}{2(\sigma_1^2+\sigma_2^2)}\right\} \\
= & \phi\left(\frac{\alpha_1\mu-(\frac{\sigma_2}{\sigma_1}x+\frac{\sigma_1}{\sigma_2}y)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)\phi\left(\frac{x-y}{\sqrt{\sigma_1^2+\sigma_2^2}}\right),
\end{aligned}$$

where $\alpha_1 = \dfrac{\sigma_1}{\sigma_2}+\dfrac{\sigma_2}{\sigma_1}$.

Hence,

$$\begin{aligned}
& \frac{1}{\sigma_1}\phi\left(\frac{x-\mu}{\sigma_1}\right)\frac{1}{\sigma_2}\phi\left(\frac{y-\mu}{\sigma_2}\right) \\
= & \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\phi\left(\frac{\alpha_1\mu-(\frac{\sigma_2}{\sigma_1}x+\frac{\sigma_1}{\sigma_2}y)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\phi\left(\frac{x-y}{\sqrt{\sigma_1^2+\sigma_2^2}}\right). \quad (3.13)
\end{aligned}$$

Using the identity $v-\dfrac{av+bw}{a+b} = \frac{b}{a+b}(v-w)$, we can write

$$\begin{aligned}
x-y &= \frac{x-\frac{\sigma_2^2x+\sigma_1^2y}{\sigma_1^2+\sigma_2^2}}{\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}} \\
&= \frac{x-(\frac{\sigma_2}{\sigma_1}x+\frac{\sigma_1}{\sigma_2}y)/\alpha_1}{\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}}.
\end{aligned}$$

Thus, using this result, we can re-write the second density in expression (3.13) to

give

$$\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\phi\left(\frac{\alpha_1\mu-(\frac{\sigma_2}{\sigma_1}x+\frac{\sigma_1}{\sigma_2}y)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\phi\left(\frac{x-(\frac{\sigma_2}{\sigma_1}x+\frac{\sigma_1}{\sigma_2}y)/\alpha_1}{\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\sqrt{\sigma_1^2+\sigma_2^2}}\right)\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2}$$

$$= \phi\left(\frac{\alpha_1\mu-(\frac{\sigma_2}{\sigma_1}x+\frac{\sigma_1}{\sigma_2}y)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)\frac{1}{\sigma_1^2}\phi\left(\frac{x-(\frac{\sigma_2}{\sigma_1}x+\frac{\sigma_1}{\sigma_2}y)/\alpha_1}{\frac{\sigma_1^2}{\sigma_2^2}\alpha_2}\right), \tag{3.14}$$

where $\alpha_2 = \dfrac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}$.

3.4 Statistical properties of an estimator

Point estimation involves inference about the unknown parameter θ. A random quantity $\hat{\theta}$, defined as a statistic or estimator for θ, is obtained as a function of the random sample $X_1, ..., X_n$. We want this statistic to be as 'close' to the true value as possible and there are several ways in which we can measure the reliability of this statistic.

3.4.1 Bias and mean squared error

The following two properties are fundamental for assessing the accuracy and precision of an estimator.

Definition 3.4.1 (Bias)**.** Suppose $\hat{\theta}$ is an estimator of θ. Then the bias of $\hat{\theta}$ is defined by

$$b(\hat{\theta}) = \mathrm{E}[\hat{\theta}] - \theta.$$

Note that in general, bias is a function of θ.

Definition 3.4.2 (Mean squared error)**.** The mean squared error (MSE) of $\hat{\theta}$ is given by

$$MSE(\hat{\theta}) = \mathrm{E}[(\hat{\theta}-\theta)^2] = \mathrm{var}(\hat{\theta}) + b^2(\hat{\theta}).$$

By Chebyshev's inequality [Lindgren, 1993] it can be seen that for every $\epsilon > 0$,

$$P(|\hat{\theta} - \theta| > \epsilon) \leq \frac{1}{\epsilon^2} E[(\hat{\theta} - \theta)^2].$$

This indicates that if the mean squared error is small, then the chance that $\hat{\theta}$ is far from θ is also small. Furthermore, if an estimator has zero bias then its MSE is its variance. Hence, the MSE is a useful measure of the error of estimation, where a desired estimator would be one with a small MSE.

3.4.2 Sufficiency and completeness

This section provides statistical concepts that are referred to throughout this thesis.

Definition 3.4.3 (Sufficiency)**.** Suppose $X_1, ..., X_n$ is a random sample from X that has density function $f_\theta(x)$. Then a statistic $T(X_1, ..., X_n)$ is a *sufficient statistic* for θ if the conditional distribution of $X_1, ..., X_n$ given $T(X_1, ..., X_n)$ is not a function of θ.

Definition 3.4.4 (Minimal sufficiency)**.** If $T(X_1, ..., X_n) = f_\theta(\tilde{T}(X_1, ..., X_n))$ for every sufficient statistic $\tilde{T}(X_1, ..., X_n)$, then $T(X_1, ..., X_n)$ is *minimal sufficient.*

Definition 3.4.5 (Completeness)**.** A statistic $T(X_1, ..., X_n)$ is *complete* if, for any function $g_\theta(T(x_1, ..., x_n))$, $E[g_\theta(T(x_1, ..., x_n))] = 0$ for all θ implies that $P_\theta(g(T(x_1, ..., x_n)) = 0) = 1$ for all θ.

Theorem 3.4.6 (Factorisation Criterion)**.** *A statistic $T(X_1, ..., X_n)$ is a sufficient statistic for θ if there exists non-negative functions g and h such that,*

$$f_\theta(x_1, ..., x_n) = g_\theta(T(x_1, ..., x_n))h(x_1, ..., x_n) \quad \forall\theta.$$

The following theorem and definition are useful for finding minimal sufficient statis-

tics.

Theorem 3.4.7 (Lehmann-Scheffé)**.** *Suppose $X_1, ..., X_n$ is a random sample from X that has density function $f_\theta(x)$ and $T(X_1, ..., X_n)$ is both sufficient and complete for θ. Then $T(X_1, ..., X_n)$ is the minimal sufficient statistic.*

Definition 3.4.8 (Complete minimal sufficiency of exponential families)**.** Suppose $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ is a random sample from $\boldsymbol{X}$ which has a $p-$variate distribution with parameter vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$, where the pdf of $\boldsymbol{X}$ is given by expression (3.7). Then the statistic

$$T(\boldsymbol{X}_1, ..., \boldsymbol{X}_n) = \left(\sum_{i=1}^{n} t_1(X_i), ..., \sum_{i=1}^{n} t_p(X_i) \right) \tag{3.15}$$

is a complete, minimal sufficient statistic for $\boldsymbol{\theta}$.

If one-to-one transformations of $T(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ exist then they are also complete sufficient statistics, and are minimal sufficient if they are $p-$dimensional. An example of this was given in Section 3.3.3.

In summary, a sufficient statistic can be found by the Factorisation Theorem. A minimal sufficient statistic can be found through either the Lehmann-Scheffé Theorem or by the theory of exponential families. These theorems combined are used in the process of Rao-Blackwellisation, as described in the next section.

3.5 Minimum variance unbiased estimation

In point estimation, bias and efficiency of estimators are two properties considered when searching for the best estimator. More than one unbiased estimator may exist, and therefore criteria by which estimators may be compared for efficiency need to be considered.

Definition 3.5.1 (Unbiasedness)**.** An estimator $\hat{\theta}$ for θ is unbiased if

$$\mathrm{E}[\hat{\theta}] = \theta \quad \forall\theta.$$

Definition 3.5.2 (Efficiency)**.** Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for θ. If $MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$ then $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ and thus a more precise estimator.

A desirable estimator for θ is therefore one which is both unbiased and has smallest variance among all unbiased estimators. This is formally known as the Uniformly Minimum Variance Unbiased Estimator (UMVUE). Utilising the concepts of Sufficiency (Definition 3.4.3) and Completeness (Definition 3.4.5), the following theorem is used for deriving the UMVUE. However, it is important to note that as an unbiased estimator may not necessarily exist, a UMVUE may not always exist.

Theorem 3.5.3 (Rao-Blackwell)**.** *If T is a sufficient statistic for θ and $\hat{\theta}$ is an unbiased estimator of θ, then $\tilde{\theta} = \mathrm{E}[\hat{\theta}|T]$ is an unbiased estimator of θ with variance at most the variance of $\hat{\theta}$, that is, $var(\tilde{\theta}) \leq var(\hat{\theta})$.*

Proof. Unbiasedness follows from

$$\mathrm{E}[\tilde{\theta}] = \mathrm{E}[\mathrm{E}[\hat{\theta}|T]] = \mathrm{E}[\hat{\theta}] = \theta$$

and for the variance it follows that

$$\begin{aligned} var(\hat{\theta}) &= \mathrm{E}[(\hat{\theta} - \mathrm{E}[\hat{\theta}])^2] \\ &= \mathrm{E}[(\hat{\theta} - \tilde{\theta})^2] + \mathrm{E}[(\tilde{\theta} - \mathrm{E}[\tilde{\theta}])^2] \\ &= \mathrm{E}[(\hat{\theta} - \tilde{\theta})^2] + var(\tilde{\theta}) \\ &\geq var(\tilde{\theta}). \end{aligned}$$

□

This is known as variance reduction or *Rao-Blackwellisation* and shows that $\tilde{\theta}$ is an unbiased and efficient estimator for θ. If T is a complete, sufficient statistic, that is, if it is a minimal sufficient statistic, then $\tilde{\theta}$ is the UMVUE for θ [Lehmann and Scheffé, 1950].

Note, although a UMVUE has smallest variance among all unbiased estimators, biased estimators may still exist with smaller MSE.

Chapter 4

Time-to-event data analysis

4.1 Introduction

Time-to-event data may arise in various areas of research, from finance to engineering to medical research, where interest may be to model the lifetime of an electronic industrial component or the survival time of a patient in a clinical trial. Therefore, depending on the context, over the past few decades, developments have been made under various names, all with the same aim of modelling the time until the event of interest occurs. These advancements have led to a large number of textbooks on the topic, including Collett [2015], Kleinbaum and Klein [2012] and Machin et al. [2006]. Modelling of time-to-event outcomes may vary from standard parametric models such as Weibull survival times to more flexible parametric and semi-parametric models such as the Cox proportional hazards model. As the focus of this thesis is clinical trials, examples of time-to-event data will come from medical research and the terms time-to-event and survival are used interchangeably.

This chapter introduces standard single stage statistical methods for the analysis of time-to-event data. The methods described in subsequent sections are based on the

textbooks mentioned above.

4.2 Survival analysis

Survival analysis is used to analyse data where the endpoint is a measure of time to an event. The endpoint is typically referred to as the event time or survival time, where time is measured from the time of origin, usually the time of entry into a trial, until the time the event occurs. Although the term survival analysis indicates death is the event of interest, other outcomes such as time to disease progression or time to tumour recurrence may be analysed using the same methods. Regardless of the type of event, the methods used are often referred to as survival analysis in medical research.

Trials with survival time-to-event outcomes tend to be longer in duration as the event of interest, such as death, typically takes longer to be observed. This is commonly defined as Overall Survival (OS) in late phase oncology trials. Such an outcome has a lower event rate and therefore a longer observation period is needed such that a sufficient number of deaths are observed for adequate power.

Although this chapter focusses on single stage methods, for two-stage trials, a plausible short-term outcome may exist with a higher event rate, which may be considered as a surrogate endpoint for the primary outcome measure at the interim analysis. Progression Free Survival (PFS) is defined as the time from entry into a trial until disease progression or death. Therefore, PFS has a higher event rate than OS and is thus commonly considered a surrogate endpoint for OS. Utilising a short-term surrogate endpoint may be desirable in cases with a long follow-up time, where the time required to observe the expected events for OS at the interim analysis would be operationally infeasible in terms of trial duration [Di Scala and Glimm, 2011].

Due to the time-to-event nature of the endpoint, the distribution of survival times

tends to be highly skewed. This can be seen on a histogram plot of the survival times, where a small proportion of events are observed in the right tail of the plot. Therefore, time-to-event data are not typically symmetrically distributed but instead may be positively skewed. Additionally, survival times may be censored if the event of interest is not observed. This is described in the following section. Hence, standard distributions available such as the normal distribution would not be appropriate for such skewed and censored data. A key feature of survival analysis is the ability to deal with censoring that is common in time-to-event data.

4.2.1 Censoring

If the event of interest is not observed, the survival time is said to be censored. Censoring may occur for two reasons, either due to limited trial duration or if a patient has been lost to follow-up.

Let T and C denote the continuous non-negative random variables associated with survival and censoring times and t and c represent the observed survival and censoring times, respectively. There are three types of censoring encountered in survival analysis; namely, left censoring, right censoring and interval censoring. Left censoring and interval censoring are not often observed in medical research and are therefore not discussed in this section.

Right censoring occurs in survival studies where the true survival time exceeds the observed survival time. For example, at the end of study or if a patient withdraws from the trial, the event of interest occurs to the right of the point last known to be event free. In this case, t remains unknown with $t > c$. The censoring time distribution is assumed to be stochastically independent of the event time distribution. This is an important requirement in survival analysis as it means the reason for censoring is unrelated to the event of interest, and therefore any censoring

is referred to as non-informative. Since right censoring is most commonly observed in clinical trials with time-to-event data, the type of data considered in this thesis is subject to right censoring.

4.2.2 Functions in survival data

The functions described in this section are based on Collett [2015] and for ease of interpretation the event of interest is assumed to be death. The two core functions in survival analysis are the **survivor function**, $S(t)$, and the **hazard function**, $h(t)$. The survivor function measures the proportion of patients who are event free at time t and therefore gives the probability of an individual surviving to time t. Recall from Chapter 3, the probability density and probability distribution functions are denoted by $f(\cdot)$ and $F(\cdot)$, respectively. Then the survival function is given by

$$S(t) = 1 - F(t) = P(T \geq t).$$

The hazard function is a conditional density given that an individual was event free up until time t. This is commonly referred to as the *instantaneous death rate* and is defined by

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \tag{4.1}$$

The cumulative hazard function is given by

$$H(t) = \int_0^t h(u)\,\mathrm{d}u.$$

All of the above functions can be derived from one another using standard theory of conditional probabilities as follows. The conditional probability in equation (4.1)

can be written as

$$\frac{P(t \leq T < t + \Delta t)}{P(T \geq t)} = \frac{F(t + \Delta t) - F(t)}{S(t)}.$$

Therefore, the hazard function can be re-expressed as

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)} \\ &= \frac{f(t)}{S(t)} \\ &= -\frac{\mathrm{d}}{\mathrm{d}t} \log S(t) \end{aligned}$$

since $\frac{\mathrm{d}S}{\mathrm{d}t} = -f(t)$. This implies that

$$S(t) = e^{-H(t)}$$

and hence, inversely, the cumulative hazard function may be obtained from the survivor function as

$$H(t) = -\log S(t). \tag{4.2}$$

4.2.3 Kaplan-Meier estimate of the survivor function

Generally, non-parametric methods are used in life sciences to estimate the survivor function and hazard function as they do not require the form of the probability density function to be specified, and therefore, make no assumptions about the underlying distribution of survival times. An empirical estimate of the survivor function may be found by the Kaplan-Meier (K-M) method, which estimates survival probabilities at each event time.

Assume there are r distinct event times that are ordered $t_{(1)} < t_{(2)} < ... < t_{(r)}$, and without loss of generality let t_j denote the ordered event time $t_{(j)}$ $(j = 1, ..., r)$. Let

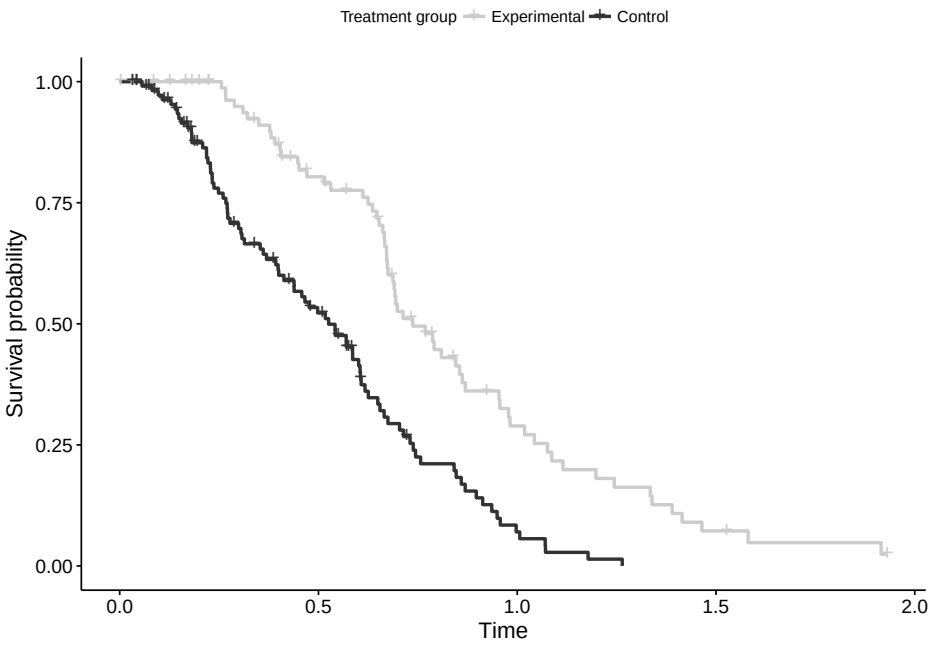

Figure 4.1: Kaplan-Meier estimate of simulated Weibull survival times for two treatment groups. Crosses indicate censored observations.

n_j denote the number of patients at risk at time t_j and d_j denote the number of patients who experience the event at time t_j. Additionally, let $R(t_j)$ denote the set of individuals alive and uncensored at a time prior to the j^{th} ordered event time t_j. The probability of surviving through an interval $[t_j, t_{j+1})$ is then estimated by $\frac{(n_j - d_j)}{n_j}$. If censored observations occur in this interval, they are assumed to occur immediately after the event time t_j so that censored patients are included in the risk set, $R(t_j)$. Since events are independent for each individual, the Kaplan-Meier estimator of the survival function is simply the product of the survival probabilities, given by

$$\hat{S}(t) = \prod_{j=1}^{l} \frac{n_j - d_j}{n_j}$$

for $t \in [t_l, t_{l+1})$ $(l = 1, ..., r)$, where $t_r = \infty$. The probability of survival in the interval t_j to $t_{j+1} - \epsilon$, for small ϵ, is unity, so that a plot of $\hat{S}(t)$ against study

time resembles a step function, as shown in the K-M plot in Figure 4.1. This is a useful tool, which is frequently used to summarise and visualise survival curves for a descriptive comparison of groups.

4.3 Hypothesis testing

4.3.1 Log-rank test

The K-M curve is an empirical estimate of the survival function which allows comparison of survival rates at a fixed time point. However, one of the most widely used statistical tests for comparing survival distributions over the whole study period is the non-parametric log-rank test.

Consider a two-arm trial with experimental treatment denoted by $k = 1$ and control treatment denoted by $k = 0$. Suppose we are interested in inference on θ, the log hazard ratio of the experimental treatment compared to control. Let n_{jk} denote the number of patients at risk at time t_j in group k and n_j denote the total number of patients at risk at time t_j. Additionally, let d_{jk} denote the total number of events observed at time t_j in group k and d_j denote the total number of events from treatment groups k at time t_j.

Summary data observed in each group at the j^{th} death time are given in Table 4.1. Under the null hypothesis of no difference in survival between treatment groups, it is known that conditional on event times, number of patients at risk at these times and the number of events at these times, d_{jk} has a hypergeometric distribution $Hyp(n_j, n_{jk}, d_j)$. Hence, by the marginal totals in Table 4.1 and by expression (3.2), the expected number of deaths in group k at time t_j is

$$e_{jk} = \frac{n_{jk} d_j}{n_j}.$$

Table 4.1: Data in each group at the j^{th} death time

Treatment group	**Observed number of deaths**	**Number alive**	**Number at risk**
Control	d_{j0}	$n_{j0} - d_{j0}$	n_{j0}
Experimental	d_{j1}	$n_{j1} - d_{j1}$	n_{j1}
Total	d_j	$n_j - d_j$	n_j

For convenience, let the proportion of patients at risk that are in group k at time t_j be defined by

$$p_{jk} = \frac{n_{jk}}{n_j}. \tag{4.3}$$

Then since d_{jk} has a hypergeometric distribution and we assume distinct death times, that is, there are no tied events, following Section 3.2, the variance of d_{jk} can be simplified to

$$v_{jk} = p_{jk}(1 - p_{jk}). \tag{4.4}$$

The score statistic is the difference in observed and expected number of deaths summed over all distinct death times. For the experimental treatment group, this is given by

$$S = \sum_{j=1}^{r} (d_{j1} - e_{j1}). \tag{4.5}$$

Under the null hypothesis, S is approximately normally distributed with mean zero and variance

$$V = \sum_{j=1}^{r} v_{j1}. \tag{4.6}$$

The square of this test statistic, divided by its variance, is thus chi-square distributed, which gives the form most commonly known as the log-rank test statis-

tic:

$$LR = \frac{S^2}{V} \sim \chi^2_1. \tag{4.7}$$

This statistic is commonly used for testing differences in survival distributions of two treatment groups as it allows for censored observations. Assuming the hazards in each group are proportional and if the number of events is not too small, S is approximately normally distributed with mean θV and variance V [Whitehead, 1997]. The proportionality of hazards is a core assumption in many survival analysis methods and is discussed in Section 4.4.1.

4.3.2 Wilcoxon test

There are several variations of the log-rank test, one of which is the Wilcoxon test, also known as the Breslow test. The Wilcoxon test statistic is given by

$$S_w = \sum_{j=1}^{r} n_j (d_{jk} - e_{jk})$$

with variance equal to $\sum_{j=1}^{r} n_j^2 v_{jk}$.

This test is similar to the log-rank test in that it compares the survivor functions between two groups. However, it incorporates a weight of the total number of patients at risk on the difference between observed and expected events. The factor of n_j ensures greater weight is assigned to early event times and less weight at times when only a small number of patients remain in the risk set. Hence, this test is more powerful when earlier differences are larger than late differences, since more weight is applied to early event times.

In addition to the Wilcoxon test, other variations of weights may be used, such as weighting by the proportion of the risk set [Tarone and Ware, 1977] or an estimate of the survivor function [Peto and Peto, 1972].

4.3.3 Stratified log-rank test

In cases where survival is expected to differ between subgroups of patients, it may be desirable to perform a stratified log-rank test to control for the subgroup differences. For example, in biomarker studies, a subpopulation with a positive biomarker may be expected to respond better to treatment compared to patients with a negative biomarker. More details on this setting are given in Chapter 7. In this case, the log-rank test statistic is calculated for each subgroup separately and then summed over all subgroups to give a stratified log-rank test statistic.

Let m denote the stratum over s strata, then the stratified log-rank statistic for comparison of two groups is given by

$$LR_{strat} = \frac{\left(\sum_{m=1}^{s} S_m\right)^2}{\sum_{m=1}^{s} V_m} \sim \chi_1^2, \tag{4.8}$$

where S_m and V_m are the log-rank test statistic and variance computed for the experimental treatment in each stratum m, as defined in equations (4.5) and (4.6), respectively.

This is a more powerful test statistic when considering subgroups as it compares the distributions of survival times in each stratum. Hence, provides a more precise summary of treatment efficacy if there are true differences between the strata in terms of hazard rate.

4.4 Estimation of the log hazard ratio

In survival analysis, the most commonly reported statistic for evaluating treatment efficacy is the hazard ratio (HR). Recall, in this thesis the log HR is denoted by θ. The methods described below for calculation of the log HR assume the hazard rates in each group are proportional over time t.

Let $\lambda_0(t)$ and $\lambda_1(t)$ denote the true hazard rates over time in the control and experimental treatment group. Then the log HR is defined by

$$\theta = \log\left(\frac{\lambda_1(t)}{\lambda_0(t)}\right).$$

The log-rank test statistic given in expression (4.7) may be used to obtain an estimate for θ by

$$\hat{\theta} = \frac{S}{V}. \tag{4.9}$$

For a large number of events, this is approximately normally distributed with mean θ and variance $\frac{1}{V}$.

Alternatively, an estimate of θ may be found by the quantities in Table 4.1. Let $O_k = \sum_{j=1}^{r} d_{jk}$ and $E_k = \sum_{j=1}^{r} e_{jk}$ denote the total observed and expected number of events for each group k ($k = 0, 1$). Then an estimate of the log HR is given by

$$\hat{\theta} = \log\left(\frac{O_1/E_1}{O_0/E_0}\right). \tag{4.10}$$

This gives an estimate of the relative death rates in the two groups. The log HR is assumed to be constant over study follow-up and therefore assumes the hazard functions are proportional over time. Under the null hypothesis of no survival difference between the two groups, the log HR is equal to 0. Hence, if $\hat{\theta}$ is less than 0, the hazard in the control group is larger than the hazard in the experimental treatment group, which therefore indicates a beneficial treatment effect.

An adjusted log HR estimate may be obtained in a similar way for the stratified log-rank test (4.8). As before, estimates for O_{km} and E_{km} are obtained separately for each treatment k in each stratum m and then summed over s strata to give a

stratified log HR. This is given by

$$\hat{\theta}_{strat} = \log\left(\frac{\sum_{m=1}^{s} O_{km}/\sum_{m=1}^{s} E_{km}}{\sum_{m=1}^{s} O_{cm}/\sum_{m=1}^{s} E_{cm}}\right). \tag{4.11}$$

4.4.1 Proportional hazards assumption

Most survival analysis methods assume hazard rates are proportional over time for all model covariates. However, this assumption may not always hold, for example, if the experimental treatment is found to be better in terms of short term survival while the control treatment is better in terms of long term survival. In this case, the hazard rates for each treatment depend on time, and thus do not satisfy the proportionality assumption. Modelling approaches described in the next section assume proportional hazards and therefore it is important to check this assumption holds before modelling.

A simple graphical check of proportionality may be to plot the log of the cumulative hazard function (4.2) against log of time, that is, $\log(-\log(S(t)))$ vs. $\log(t)$. If the proportional hazards assumption holds, the lines should appear approximately parallel. Additionally, time-dependent covariates may be incorporated to check for proportionality when modelling.

4.5 Modelling approaches

Most modelling principles of survival data are similar to those for other types of data, except survival data are generally modelled on the log-hazard scale. The two most common approaches are either parametric models or semi-parametric models. The former assumes some functional form for the baseline hazard, whereas the latter makes no assumption about the shape of the baseline hazard function. This relaxed assumption gives the most widely adopted model in survival analysis known as the

Cox proportional hazards model.

4.5.1 Cox proportional hazards model

The most common model applied to time-to-event data is the semi-parametric Cox proportional hazards (PH) model [Cox, 1972]. For an individual i ($i = 1, ..., n$), let $\boldsymbol{x}_i = (x_{i1}, ..., x_{ig})$ denote a vector of g predictor variables and $\boldsymbol{\beta} = (\beta_1, ..., \beta_g)'$ denote the associated vector of coefficients. Then the hazard function for the i^{th} individual can be modelled by

$$h_i(t) = h_0(t)e^{\boldsymbol{\beta}'\boldsymbol{x}_i}, \tag{4.12}$$

where $h_0(t)$ is the baseline hazard function and $\boldsymbol{\beta}'\boldsymbol{x}_i$ is the linear predictor, also known as the risk score or prognostic index for the i^{th} individual. The baseline hazard function is the hazard function for an individual with all covariate values $\boldsymbol{x} = (0, ..., 0)$. Thus an estimate of the log HR ($\hat{\theta}$) for covariates $\boldsymbol{x}_i$ relative to $\boldsymbol{x} = 0$ is $\hat{\boldsymbol{\beta}}$.

The form of this model indicates that the covariates act multiplicatively on the baseline hazard rate at any point in time and therefore implies the hazard functions for all variables are proportional over time.

The most appealing feature of the Cox model is that it does not estimate the baseline hazard function and therefore makes no assumption about the probability distribution of survival times. Hence, a semi-parametric approach is taken for parameter estimation, where the partial likelihood is maximised to estimate the $\boldsymbol{\beta}$ parameters.

Let δ_i denote the indicator for an uncensored survival time at t_i and let $R(t_i)$ denote the risk set at time t_i for the i^{th} individual. Then the partial likelihood function is

given by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ \frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{\sum_{q \in R(t_i)} \exp(\boldsymbol{\beta}'\boldsymbol{x}_q)} \right\}^{\delta_i}.$$

4.5.2 Parametric models

In some cases it may be reasonable to assume the baseline hazard follows a specific probability distribution. In such cases, a fully parametric survival model may be fitted to give more precise estimates. The most common distributions considered in survival analysis are the exponential and Weibull distributions.

The simplest case is when it is assumed that the hazard is constant over time. In this case the hazard function can be modelled by $h_0(t) = \lambda$ for $0 \leq t < \infty$. This therefore gives the exponential survivor function $S(t) = e^{-\lambda t}$ with probability density function of survival times $f(t) = \lambda e^{-\lambda t}$. Thus, the exponential PH model can be written as

$$h_i(t) = \lambda e^{\boldsymbol{\beta}'\boldsymbol{x}_i}. \tag{4.13}$$

Due to the constant hazard function, the exponential distribution has a 'lack of memory' property such that

$$p(T > t) = p(T > t + t_0 | T > t_0)$$

for $t_0 > 0$. Hence, the probability of surviving another t period is independent of surviving up until time t_0. This is an attractive feature of the exponential model, however it is not often representative of medical data. In medical research it is typical for the baseline hazard to vary with time, and therefore the constant hazard function of the exponential model is not appropriate. A generalisation of this model gives the most popular parametric survival model known as the Weibull model. This

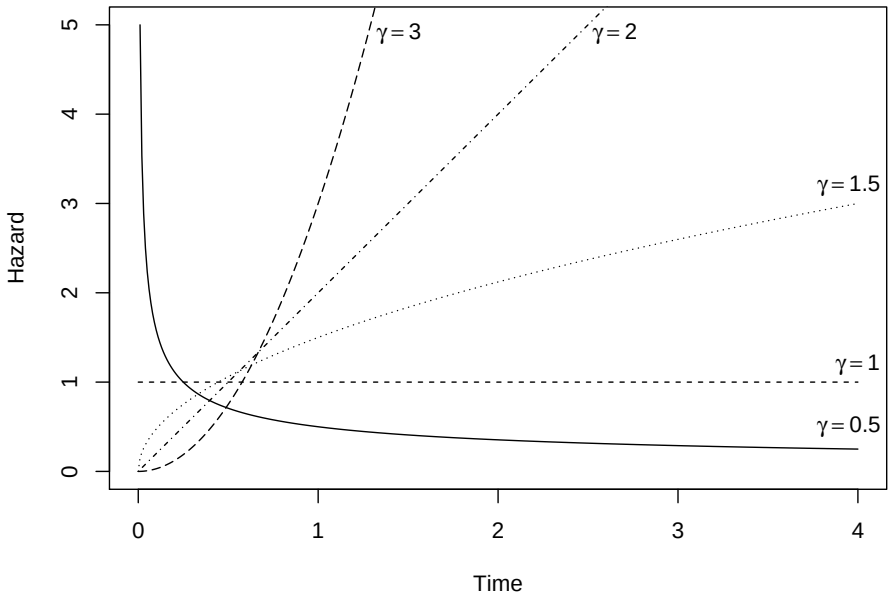

Figure 4.2: Weibull hazard function for $\gamma \in \{0.5, 1, 1.5, 2, 3\}$ with $\lambda = 1$.

allows the hazard function to vary with time as a power of t and thus provides a more flexible form of the hazard function. This is given by $h_0(t) = \lambda\gamma t^{\gamma-1}$. Assuming survival times follow a Weibull distribution, the probability density function is $f(t) = \lambda\gamma t^{\gamma-1}e^{-\lambda t^{\gamma}}$ with survivor function $S(t) = e^{-\lambda t^{\gamma}}$. Hence, the Weibull PH model is given by

$$h_i(t) = \lambda\gamma t^{\gamma-1}e^{\boldsymbol{\beta}'\boldsymbol{x}_i}. \tag{4.14}$$

The shape parameter γ allows the hazard function to vary with time. Hence, when $\gamma = 1$, the Weibull model reverts to the exponential model. Examples of Weibull hazard functions for different values of γ and $\lambda = 1$ can be seen in Figure 4.2. For $\gamma > 1$ the hazard rate increases as γ increases, while for $\gamma < 1$, the hazard decreases.

Since these models are fully parametric, model parameters, λ and γ, are estimated

by maximising the likelihood function

$$\prod_{i=1}^{n} \{f(t_i)\}^{\delta_i} \{S(t_i)\}^{1-\delta_i}.$$

The multiplicative form of these models again implies the proportional hazards assumption must be satisfied for model validity. For data with non-proportional hazards, these models are not appropriate as the assumption of a constant hazard ratio over time is violated. In this case, there are several more complex methods which may be used that incorporate time-dependent covariates, such as stratified Cox regression or flexible parametric modelling [Royston and Lambert, 2011].

In addition to the models presented in this section, other distributions may be considered such as the Gompertz, Gamma and Log-Normal distribution [Collett, 2015].

Chapter 5

Unbiased estimation in seamless phase II/III clinical trials with normally distributed outcomes

This chapter introduces methods of estimation for two-stage trials with normally distributed data that we will build upon in the next chapter. The first section describes issues that arise in estimation when combining data across two stages. Specifically, selection bias of naive maximum likelihood estimators is explored for various trial scenarios. Previously suggested unbiased and efficient estimators are then described for normally distributed outcomes. The chapter ends with a simulation study that compares the properties of the naive and unbiased estimators.

5.1 Setting

This chapter assumes the setting of a phase II/III clinical trial described in Section 2.4.1. As discussed, combining data from stages 1 and 2 may result in biased estimation of the treatment effect size, as stage 1 data are used for both treatment selection as well as testing for treatment efficacy. This is because selection is made on the basis of ordered effect sizes at the interim analysis, where the most efficacious treatment is selected to take forward into stage 2. Therefore, early selection may result in an overestimated treatment effect at the end of the trial. Using stage 2 data alone will yield an unbiased estimate of the treatment effect, but at the expense of losing information from stage 1. Thus, an ideal estimator would be one which utilises all trial data while correcting for the potential bias due to treatment selection.

The degree of selection bias varies depending on several factors such as the planned sample size, the proportion of patients in each stage, the number of treatments under investigation, and the true effect sizes for the selected and dropped treatments [Bauer et al., 2010]. The influence of these factors on selection bias are discussed in the following sections.

5.2 Notation

We use the notation in Bowden and Glimm [2008] for a two-stage phase II/III clinical trial. Assume treatment arms in stage 1 are labelled $k = 0, 1, ..., K$, where $k = 0$ corresponds to the control treatment, with outcomes for each treatment normally distributed with unknown mean μ_k and known common variance σ^2. Let n_i denote the number of patients allocated to each treatment arm in stage $i = 1, 2$. Let $\bar{X}_k$ denote the stage 1 sample means for treatment k such that $\bar{X}_k \sim N(\mu_k, \sigma_1^2)$, where

$\sigma_1^2 = \sigma^2/n_1$. At the interim analysis, the treatment deemed most efficacious based on a predefined criterion is selected, along with the control treatment, for further analysis in stage 2. Let s ($s \in \{1, ..., K\}$) denote the index of the selected treatment such that stage 2 sample means are denoted by $\bar{Y}_j \sim N(\mu_j, \sigma_2^2)$ for $j \in \{0, s\}$, where $\sigma_2^2 = \sigma^2/n_2$. Let τ denote the information fraction available at the interim analysis, which is defined by the ratio of the stage 1 sample size per group to the total sample size per group, that is, $\tau = n_1/n$ where $n = n_1 + n_2$. This is also referred to as the 'selection time'.

5.3 Selection rule

At the end of stage 1, various interim adaptations may be applied before the trial continues to stage 2. As mentioned, these may include, stopping early for futility or efficacy, modifying the total sample size, or selecting the best experimental treatment. In this section, based on Kimani et al. [2013], we explore the impact of treatment selection, where the treatment deemed most effective at the interim analysis is selected to continue in stage 2. For normally distributed outcomes, sample means are used to assess treatment efficacy. At the interim analysis, stage 1 sample means are ordered in increasing magnitude where the treatment with the maximum observed mean is selected to continue in stage 2. Let $\bar{X}_{(1)} > \bar{X}_{(2)} > ... > \bar{X}_{(K)}$ denote the ordered stage 1 sample means, then the selected treatment s is such that $\bar{X}_s = \bar{X}_{(1)}$.

5.4 Maximum likelihood estimators

At the final analysis, interest lies in estimating the mean treatment difference $\theta_s = \mu_s - \mu_0$ whilst utilising data from both stages of the trial. We can obtain a naive

Maximum Likelihood Estimate (MLE) for each treatment group separately. This is the weighted average of stage 1 and stage 2 data, which ignores treatment selection. Let $\hat{\mu}_k$ denote the naive MLE for μ_k, $k = \{0, s\}$, such that

$$\hat{\mu}_0 = \tau \bar{X}_0 + (1 - \tau)\bar{Y}_0 \tag{5.1}$$

and

$$\hat{\mu}_s = \tau \bar{X}_s + (1 - \tau)\bar{Y}_s. \tag{5.2}$$

Then the naive MLE for the mean treatment difference is given by

$$\hat{\theta}_s = \hat{\mu}_s - \hat{\mu}_0. \tag{5.3}$$

This is commonly referred to as a 'naive' estimator as it does not consider the adaptation rule at the interim analysis. If the only interim adaptation is that of treatment selection and there is no possibility of early stopping, then $\hat{\mu}_0$ is unbiased for the control treatment mean. On the other hand, for the selected treatment, $\hat{\mu}_s$ is biased for μ_s due to the ordering and selection of $\bar{X}_s$. Subsequently, $\hat{\theta}_s$ is biased for θ_s.

5.5 Selection bias

Selection bias is defined as the expected difference between the estimator and the true mean treatment-control difference, that is, $E[\hat{\theta}_s - \theta_s]$. Several papers discuss bias in estimation that arises from treatment selection where they have shown that selection may lead to overestimation of the true mean difference [Bauer et al., 2010, Bretz et al., 2009, Posch et al., 2005]. For example, if all experimental treatments are equally effective then selecting a treatment that has the largest mean effect at

the interim analysis will result in overestimating the mean difference at the final analysis since it is selected only by chance. Therefore, this leads to a positively biased estimator [Bauer et al., 2010].

Analytical expressions for bias and Mean Squared Error (MSE) have been given by several authors. Bauer et al. [2010] show that treatment selection bias depends on a combination of the number of experimental treatments, selection time and variance of the sample means. Let ϕ_{μ,σ^2} and Φ_{μ,σ^2} denote the density and cumulative distribution functions of a normal distribution $N(\mu, \sigma^2)$, respectively. Then the following expressions by Bauer et al. define the selection bias and MSE of the naive estimator, given by equation (5.3), when the treatment with maximum mean is selected at any point during the trial, that is, where $\tau \leq 1$.

$$\begin{aligned} Bias &= \sum_{k=1}^{K} E[\hat{\mu}_s - \mu_k | \hat{\mu}_s = \hat{\mu}_k] P(\hat{\mu}_s = \hat{\mu}_k) \\ &= \tau \sum_{k=1}^{K} \int_{-\infty}^{\infty} (x - \mu_k) \phi_{\mu_k, \sigma_1^2}(x) \Pi_{j \neq k}^{K} \Phi_{\mu_j, \sigma_1^2}(x) \mathrm{d}x \end{aligned} \tag{5.4}$$

and

$$\begin{aligned} MSE &= \sum_{j=1}^{K} E[(\hat{\mu}_s - \hat{\mu}_0 - (\mu_j - \mu_0))^2 | \hat{\mu}_s = \hat{\mu}_j] P(\hat{\mu}_s = \hat{\mu}_j) \\ &= \tau^2 \sum_{k=1}^{K} \int_{-\infty}^{\infty} (x - \mu_k)^2 \phi_{\mu_k, \sigma_1^2}(x) \prod_{j \neq k}^{K} \Phi_{\mu_j, \sigma_1^2}(x) \mathrm{d}x + (2 - \tau)\sigma^2/n. \end{aligned} \tag{5.5}$$

Expressions (5.4) and (5.5) demonstrate that bias and MSE depend on the selection time τ, mean outcomes μ_k, number of treatments K, variance σ^2 and the sample size n. Since τ is the ratio of the stage 1 sample size to the total sample size, from the contribution of τ in the above expressions, we can conclude that bias and MSE are maximum when selection is made at the final analysis, that is, when $\tau = 1$. This is commonly referred to as 'post-trial selection' [Bauer et al., 2010].

Using expressions (5.4) and (5.5), we now assess the bias and root mean squared error (RMSE) of the naive MLE for a range of values of τ and K for different scenarios of the true mean effects μ_k. For each scenario we consider values of $K \in \{1, ..., 6\}$ treatment groups. Let the selection time $0 \leq \tau \leq 1$, such that the number of patients in each treatment group in stage 1 is $n_1 = \tau n$. A total of $n = 100$ patients per group are assumed. However, to make results approximately independent of n, bias and RMSE plots are given in units of $\sigma\sqrt{2/n}$. This therefore allows generalisation of results to trials with different sample sizes in certain ranges.

5.5.1 Equal treatment means

First consider the setting where all treatment means are equal, and without loss of generality let $\mu_k = 0 \ \forall \ k$. This is considered the worst case scenario where we would expect greatest bias [Bauer et al., 2010]. Figure 5.1(a) shows the bias (top) and RMSE (bottom) for $\hat{\theta}_s$ as a function of selection time, τ, for up to 6 treatment-control comparisons. It can be seen that selection bias increases with selection time (proportion of stage 1 patients). Hence, as expected, the maximum bias is observed in the case of post-trial selection. Additionally, bias increases with the number of experimental treatment groups K.

In terms of MSE, Figure 5.1(a) (bottom) indicates an increase in RMSE as the number of treatments in a given trial increases. However, this increase is not as steep as compared with selection bias (note the difference in scales). An explanation for this is the square of the information fraction in the expression for the MSE in equation (5.5). For $K = 2$, Figure 5.1(a) (bottom) shows that RMSE is equal to that for $K = 1$. This is expected as we have assumed equal treatment means with a common variance and Posch et al. [2005] show that selection between two normally distributed outcomes with the same variance only depends on the effect size.

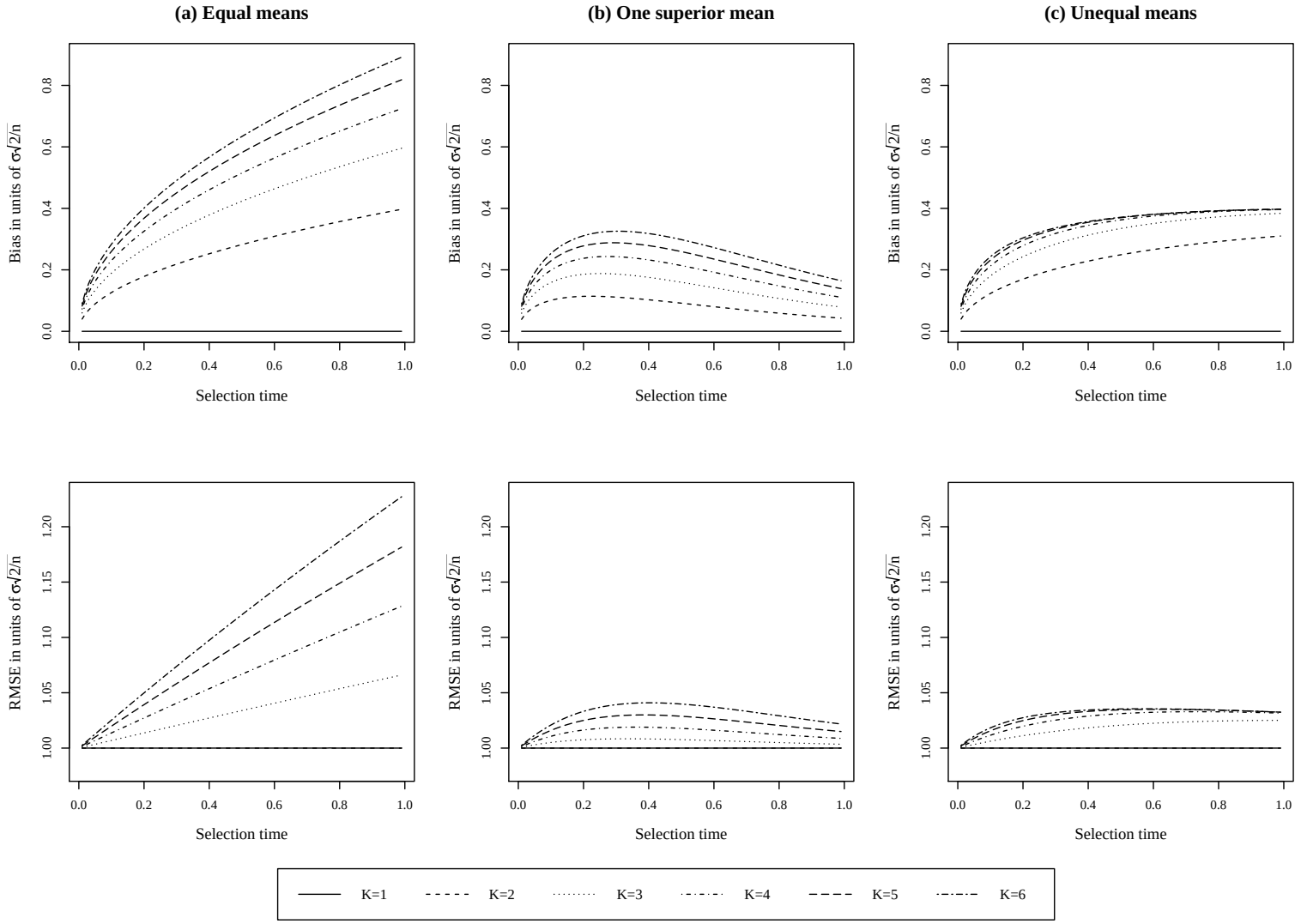

Figure 5.1: Bias (top row) and RMSE (bottom row) depending on the number of experimental treatment arms $K \in \{1, ..., 6\}$ and the selection time $0 \leq \tau \leq 1$. Each column represents a different combination of treatment means, μ_k, $k \in \{0, ..., K\}$: (a) $\mu_k = 0$, (b) $\mu_1 = 0.3$ and $\mu_{k \neq 1} = 0$, and (c) $\mu_k = 0.1k$.

5.5.2 Unequal treatment means

In the case of unequal treatment means, we assess the effect of the magnitude of the treatment difference on selection bias. We consider two scenarios to assess the bias and MSE for different values of τ and K.

In the first scenario we assume one uniquely superior experimental treatment with treatment means $\mu_1 = 0.3$ and $\mu_k = 0$ $(k = 2, ..., K)$. Table 5.1 gives the probability of selecting the treatment with the largest mean effect. For all values of K, it can be seen that if one treatment is uniquely superior to all other experimental treatments, then the probability of selecting this treatment is increased compared to the case of equal means. Hence, as the probability of selection increases, selection bias reduces. This can be seen in a comparison of Figures 5.1(a) and 5.1(b).

Furthermore, as we have seen in the previous section, the information fraction τ is an important factor for selection bias. In Table 5.1 we can see that the probability of correct selection increases as τ increases. For example, for a trial with two experimental treatment arms, the probability of correct selection increases from 0.75 to 0.91 for $\tau = 0.2$ to $\tau = 0.8$. This is because an early selection time may lead to premature data with reduced precision, which consequently reduces the chance of correctly selecting the truly superior treatment.

In the second scenario we consider unequal treatment means $\mu_k = 0.1k$ for $k = \{0, ..., K\}$. To assess the effect of τ in this scenario, Figure 5.1(c) shows a steady increase in bias and RMSE until $\tau = 0.3$, at which point bias and RMSE reach a plateau for all values of K. Furthermore, in contrast to the scenario of equal treatment means, where bias increases with K, in this scenario, bias is found to be approximately equal for all values of $K \geq 4$ for $\tau > 0.3$. This is similar for RMSE. In addition, a higher probability of selection is associated with a larger information fraction τ, while an increase in the number of experimental treatments K leads to

Table 5.1: Probability of selecting the truly most effective experimental treatment for normally distributed outcomes. Treatment arms are denoted by $k \in \{0, ..., K\}$, where $k = 0$ denotes the control arm. A range of treatment means $\mu_k \in \{0, ..., 0.6\}$ are considered for varying information fraction τ under 3 scenarios of equal effect sizes, one superior treatment, and unequal effect sizes.

		No. of treatment arms, K					
τ	Effect size	1	2	3	4	5	6
0.2	$\mu_k = 0$	1	0.50	$0.3\dot{3}$	0.25	0.20	$0.1\dot{6}$
	$\mu_1 = 0.3, \mu_{k\neq 1} = 0$	1	0.75	0.62	0.54	0.48	0.43
	$\mu_k = 0.1k$	1	0.59	0.47	0.42	0.40	0.39
0.5	$\mu_k = 0$	1	0.50	$0.3\dot{3}$	0.25	0.20	$0.1\dot{6}$
	$\mu_1 = 0.3, \mu_{k\neq 1} = 0$	1	0.86	0.77	0.70	0.65	0.61
	$\mu_k = 0.1k$	1	0.64	0.55	0.52	0.51	0.51
0.8	$\mu_k = 0$	1	0.50	$0.3\dot{3}$	0.25	0.20	$0.1\dot{6}$
	$\mu_1 = 0.3, \mu_{k\neq 1} = 0$	1	0.91	0.85	0.80	0.76	0.73
	$\mu_k = 0.1k$	1	0.67	0.60	0.58	0.58	0.58

reduced probability of selection (Table 5.1).

To further assess the effect of the magnitude of the treatment differences, we consider the case of 'post-trial selection', where $\tau = 1$ for a trial with $K = 3$ experimental treatments. Figures 5.2 and 5.3 illustrate contours of selection bias and RMSE respectively, in units of $\sigma\sqrt{2/n}$. The range of treatment means for μ_2 and μ_3 are given along the x and $y - axis$, respectively, with $\mu_0 = \mu_1 = 0$ in all configurations.

In Figure 5.2, maximal selection bias is observed at the centre of the plot where all treatment means are equal, that is, when $\mu_1 = \mu_2 = \mu_3$. As can be seen, bias reduces as the difference in means increases. Hence, this reiterates that selection bias depends on the size of the mean difference.

In terms of RMSE, for a trial with 3 experimental treatment groups, contours of RMSE form an elliptical shape where again maximal RMSE is observed at the centre where all treatment means are equal (Figure 5.3). As can be seen, RMSE decreases with increasing difference in means.

In conclusion, selection bias and MSE depend on the selection time τ, number of experimental treatments K, the variance of the effect estimates, the size of the treatment difference and the sample size.

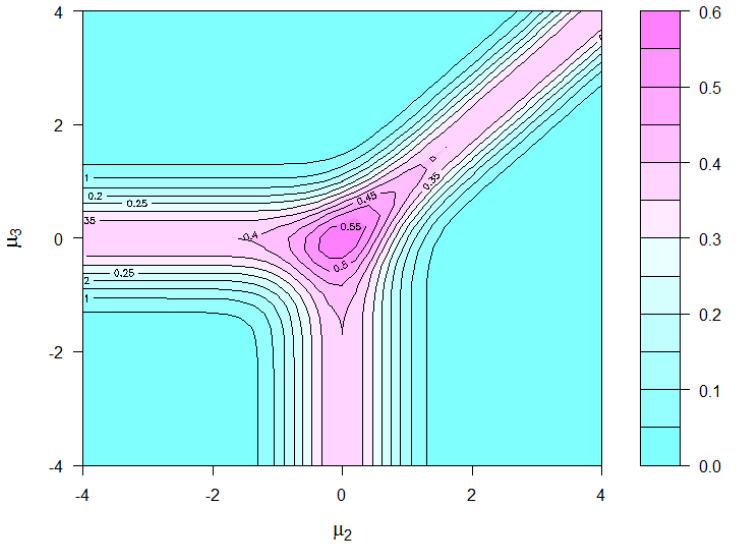

Figure 5.2: Contours of selection bias for $K = 3$ and $\tau = 1$

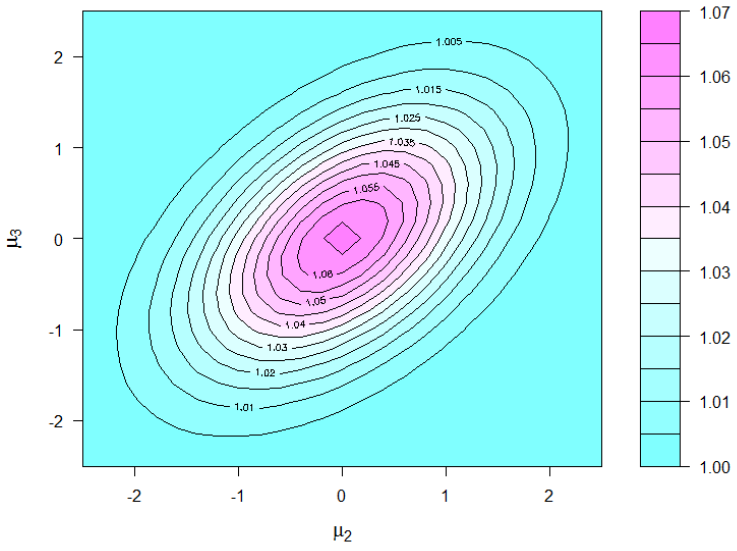

Figure 5.3: Contours of RMSE for $K = 3$ and $\tau = 1$

5.6 Unbiased estimation for normal outcomes

The MLE presented in expression (5.3) combines both stages of trial data to obtain an estimator for the selected treatment effect. As we have seen, this estimator is biased as stage 1 data are used for both treatment selection and estimation at the end of the trial. It is possible to obtain an unbiased estimator using stage 2 data alone, however the precision of this estimator would be compromised as it does not utilise all trial data. Several methods that account for selection bias in estimation, which utilise all trial data, are available for normally distributed outcomes. One such method applies the Rao-Blackwell theorem to obtain unbiased and efficient estimators conditional on various interim adaptations [Cohen and Sackrowitz, 1989, Kimani et al., 2013]. This method is now described in detail.

As demonstrated in Section 5.5, estimation is conditional on the ranking of treatment effects. Thus, when the selection rule is based on the observed effect size, the naive estimator $\hat{\theta}_s$ is positively biased for θ_s, regardless of whether the trial is able to stop early for futility. Several papers have addressed bias in estimation for a two-stage setting with normally distributed outcomes. Some focus on bias correction methods in which bias-adjusted estimators are proposed based on an iterative bias evaluation approach [Shen, 2001, Stallard and Todd, 2005], whilst others have proposed unbiased estimators for various trial settings. These include, allowing for unequal stagewise variances [Bowden and Glimm, 2008], allowing for early stopping [Kimani et al., 2013], and allowing for correlated stage 1 statistics [Robertson et al., 2016]. As this thesis is interested in unbiased estimators, the latter methods are now explored in detail with the aim of providing a basis for work in the next chapter.

Unbiased estimation in trials with treatment selection was first introduced by Cohen and Sackrowitz [1989] where they developed a Uniformly Minimum Variance Condi-

tionally Unbiased Estimator (UMVCUE) using the method of Rao-Blackwellisation, as defined by Theorem 3.5.3. They assume independent, normally distributed treatment means with equal variances and do not permit early stopping for futility or efficacy. This method has been discussed and extended by Bowden and Glimm [2008] and Kimani et al. [2013]. The former extend the Cohen and Sackrowitz [1989] estimator to consider unequal variances and extend the selection rule to not only select the most efficacious treatment but also the k^{th} best treatment out of K.

Kimani et al. [2013] further extend these methods to incorporate an early stopping futility boundary if no experimental treatment indicates sufficient efficacy at the interim analysis, whilst also conditioning on the trial continuing to the second stage. They derive a UMVCUE for μ_s and μ_0 by calculating the expected values for $\bar{Y}_s$ and $\bar{Y}_0$ conditional on sufficient and complete statistics. Hence, they give an unbiased estimator for the most efficacious experimental treatment at the interim analysis, conditional on continuing to the second stage with the possibility of stopping early for futility. This estimator is now described with the aim to derive a UMVCUE for θ_s. However, to allow comparison with estimators developed in later chapters, we consider the scenario where the trial always continues to the second stage without the possibility of early stopping.

5.6.1 A UMVCUE accounting for treatment selection

As mentioned, estimates from stage 2 data alone, $\bar{Y}_s$ and $\bar{Y}_0$, are unbiased but inefficient for μ_s and μ_0, respectively. Therefore, unbiased and efficient estimators that combine both stage 1 and 2 data are sought. Since $\bar{Y}_s$ is unbiased for μ_s, using the Rao-Blackwell theorem a new unbiased and efficient estimator is defined by $\tilde{\mu}_s = E[\bar{Y}_s|T_s]$, where T_s is a sufficient and complete statistic. As before, let stage 1 treatment means be ordered by sample means, that is, $\bar{X}_{(1)} > \bar{X}_{(2)} > ... > \bar{X}_{(k)}$ and without loss of generality, let $\bar{X}_{(k)} = \bar{X}_k$. Then the selected treatment at the

interim analysis is treatment 1 such that $s = 1$.

We are interested in the trial continuing to the second stage in order to derive the UMVCUE at the end of the trial which pools stage 1 and 2 estimates. Let $Q(X)$ denote the observed ordering of the treatment means, that is, $Q(X) = \{\bar{X}_0, \bar{X}_1 > ... > \bar{X}_k\}$. We seek the sufficient statistic T_1 given $Q(X)$ for μ_1 that combines both stage 1 and 2 means for treatment 1.

Let $I_{[Q(x)]}$ denote the indicator function for $Q(x)$ and $P(\mu) = Prob(I_{(Q(x))} = 1)$. Then the joint density of $(\bar{Y}_0, \bar{Y}_1, \bar{X})$ given $Q(X)$, where $\bar{X} = (\bar{X}_0, \bar{X}_1, ..., \bar{X}_K)$ is given by

$$f(\bar{y}_1, \bar{y}_0, \bar{x}|Q) = \frac{I_{[Q(x)]}}{P(\mu)} \frac{1}{\sigma_2}\phi\left(\frac{\bar{y}_1 - \mu_1}{\sigma_2}\right) \frac{1}{\sigma_1}\phi\left(\frac{\bar{x}_1 - \mu_1}{\sigma_1}\right) \psi(\bar{y}_0, \bar{x}'), \quad (5.6)$$

where $\bar{x}' = (\bar{x}_0, \bar{x}_2, ..., \bar{x}_K)$ and

$$\psi(\bar{y}_0, \bar{x}') = \frac{1}{\sigma_2}\phi\left(\frac{\bar{y}_0 - \mu_0}{\sigma_2}\right) \prod_{k\in\{0,2,...,K\}} \frac{1}{\sigma_1}\phi\left(\frac{\bar{x}_k - \mu_k}{\sigma_1}\right).$$

Let $\alpha_1 = \frac{\sigma_1}{\sigma_2} + \frac{\sigma_2}{\sigma_1}$ and $\alpha_2 = \frac{\sigma_2^2}{\sqrt{\sigma_1^2+\sigma_2^2}}$, then using the result in expression (3.11) (Section 3.3.5), density (5.6) can be re-written as

$f(\bar{y}_1, \bar{y}_0, \bar{x}|Q)$

$$= \frac{I_{[Q(x)]}}{P(\mu)}\phi\left(\frac{\frac{\sigma_2}{\sigma_1}\bar{x}_1 + \frac{\sigma_1}{\sigma_2}\bar{y}_1 - \mu_1\alpha_1}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) \frac{1}{\sigma_1^2}\phi\left(\frac{\bar{x}_1 - \left(\frac{\sigma_2}{\sigma_1}\bar{x}_1 + \frac{\sigma_1}{\sigma_2}\bar{y}_1\right)/\alpha_1}{\frac{\sigma_1^2}{\sigma_2^2}\alpha_2}\right) \psi(\bar{y}_0, \bar{x}). \quad (5.7)$$

Let $Z_1 = \left(\frac{\sigma_2}{\sigma_1}\bar{X}_1 + \frac{\sigma_1}{\sigma_2}\bar{Y}_1\right)$, then by Definition 3.4.3 and Definition 3.4.8, $T_1 = (\bar{X}_0, \bar{X}_2, ..., \bar{X}_K, \bar{Y}_0, Z_1)$ is a sufficient and complete statistic for estimating μ_1. Hence, by Rao-Blackwellisation, the UMVCUE for μ_1 is found by deriving $E[\bar{Y}_1|T_1, Q]$. Since $E[\bar{Y}_1|T_1, Q] = \int \bar{y}_1 f(\bar{y}_1|t_1, Q)\mathrm{d}\bar{y}_1$ and $f(\bar{y}_1|t_1, Q) = \frac{f(\bar{y}_1, t_1, Q)}{f(t_1, Q)}$, we seek the den-

sities $f(\bar{y}_1, \bar{x}_0, \bar{x}_2, ..., \bar{x}_K, \bar{y}_0, z_1|Q)$ and $f(\bar{x}_0, \bar{x}_2, ..., \bar{x}_K, \bar{y}_0, z_1|Q)$.

By transformation of random variables, the density $f(\bar{y}_1, \bar{y}_0, \bar{x}|Q)$ (5.7) is transformed into $f(\bar{y}_1, \bar{x}_0, \bar{x}_2, ..., \bar{x}_K, \bar{y}_0, z_1|Q)$ with the Jacobian of the transformation equal to $\frac{\sigma_1^2}{\sigma_2^2}$. This gives

$$f(\bar{y}_1, \bar{x}_0, \bar{x}_2, ..., \bar{x}_K, \bar{y}_0, z_1|Q) = \frac{I_{[Q(x)]}}{P(\mu)} \phi\left(\frac{z_1 - \mu_1\alpha_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) \frac{1}{\sigma_2^2} \phi\left(\frac{\bar{y}_1 - \frac{z_1}{\alpha_1}}{\alpha_2}\right) \psi(\bar{y}_0, \bar{x}'). \tag{5.8}$$

Similarly, the density $f(\bar{x}_0, \bar{x}_2, ..., \bar{x}_K, \bar{y}_0, z_1|Q)$ is derived by transforming $f(\bar{y}_1, \bar{y}_0, \bar{x}|Q)$ (5.7) into $f(\bar{x}, \bar{y}_0, z_1|Q)$, with Jacobian of the transformation equal to 1, and then integrating with respect to $\bar{x}_1$, as follows. The ordering of $Q(\bar{X})$ implies $\bar{x}_1 > \bar{x}_2$ and thus, by transformation of $f(\bar{y}_1, \bar{y}_0, \bar{x}|Q)$,

$$f(\bar{x}, \bar{y}_0, z_1|Q) = \frac{I_{[Q(x)]}}{P(\mu)} \phi\left(\frac{z_1 - \mu_1\alpha_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) \frac{1}{\sigma_1^2} \phi\left(\frac{\bar{x}_1 - \frac{z_1}{\alpha_1}}{\frac{\sigma_1^2}{\sigma_2^2}\alpha_2}\right) \psi(\bar{y}_0, \bar{x}').$$

Thus, $f(\bar{x}_0, \bar{x}_2, ..., \bar{x}_K, \bar{y}_0, z_1|Q)$

$$\begin{aligned}
&= \int_{\bar{x}_2}^{\infty} f(\bar{x}, \bar{y}_0, z_1|Q) \mathrm{d}x \\
&= \frac{I_{[Q(x)]}}{P(\mu)} \phi\left(\frac{z_1 - \mu_1\alpha_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) \frac{1}{\sigma_1^2} \left\{1 - \Phi\left(\frac{\bar{x}_2 - \frac{z_1}{\alpha_1}}{\frac{\sigma_1^2}{\sigma_2^2}\alpha_2}\right)\right\} \psi(\bar{y}_0, \bar{x}') \\
&= \frac{I_{[Q(x)]}}{P(\mu)} \frac{\alpha_2}{\sigma_2^2} \phi\left(\frac{z_1 - \mu_1\alpha_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) [1 - \Phi(-\omega_1)] \psi(\bar{y}_0, \bar{x}'),
\end{aligned} \tag{5.9}$$

where $\omega_1 = \left(\sqrt{\sigma_1^2 + \sigma_2^2}/\sigma_1^2\right)(\hat{\mu}_1 - \bar{x}_2)$. Hence, from expressions (5.8) and (5.9), the density required for the UMVCUE is

$$\begin{aligned}
f(\bar{y}_1|\bar{x}_0, \bar{x}_2, ..., \bar{x}_K, \bar{y}_0, z_1; Q) &= \frac{f(\bar{y}_1, \bar{x}_0, \bar{x}_2, ..., \bar{x}_K, \bar{y}_0, z_1|Q)}{f(\bar{x}_0, \bar{x}_2, ..., \bar{x}_K, \bar{y}_0, z_1|Q)} \\
&= \frac{\frac{1}{\alpha_2}\phi\left(\frac{\bar{y}_1 - \frac{z_1}{\alpha_1}}{\alpha 2}\right)}{1 - \Phi(-\omega_1)} I\left[\bar{y}_1 < \frac{\sigma_2}{\sigma_1}\left(z_1 - \frac{\sigma_2}{\sigma_1}\bar{x}_2\right)\right].
\end{aligned} \tag{5.10}$$

The range of $\bar{y}_1$ comes from the condition of selection $Q(\bar{X})$ where $z_1 > \frac{\sigma_2}{\sigma_1}x_2 + \frac{\sigma_1}{\sigma_2}\bar{y}_1$. Let L_1 denote this limit of integration such that $L_1 = \frac{\sigma_2}{\sigma_1}(z_1 - \frac{\sigma_2}{\sigma_1}\bar{x}_2)$. Then using the result of the expectation of a truncated normal (3.9), the UMVCUE for μ_1 is

$$\begin{aligned}
\tilde{\mu}_1 = E[\bar{Y}_1|\bar{X}_0, \bar{X}_2, ..., \bar{X}_K, \bar{Y}_0, Z_1, Q] &= \int_{-\infty}^{L_1} \bar{y}_1 f(\bar{y}_1|\bar{x}_0, \bar{x}_2, ..., \bar{x}_K, \bar{y}_0, z_1; Q)\mathrm{d}\bar{y}_1 \\
&= \frac{\frac{1}{\alpha_2}\int_{-\infty}^{L_1} \bar{y}_1 \phi\left(\frac{\bar{y}_1 - z_1/\alpha_1}{\alpha_2}\right)\mathrm{d}\bar{y}_1}{1 - \Phi(-\omega_1)} \\
&= \frac{\frac{\sigma^2}{n_2}\bar{x}_1 + \frac{\sigma^2}{n_1}\bar{y}_1}{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} - \frac{\sigma_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\frac{\phi(\omega_1)}{\Phi(\omega_1)} \\
&= \hat{\mu}_1 - \frac{\sigma_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\frac{\phi(\omega_1)}{\Phi(\omega_1)}. \qquad (5.11)
\end{aligned}$$

The first term in equation (5.11) is the naive MLE for the selected treatment $\hat{\mu}_s$ (5.2), and thus, the second term is the bias correction factor for overestimation of the MLE.

For the control treatment, the UMVCUE for μ_0 can be derived in a similar way. The density of stage 1 and 2 sample means is given by

$$f(\bar{y}_1, \bar{y}_0, \bar{x}|Q) = \frac{I_{[Q(x)]}}{P(\mu)}\frac{1}{\sigma_2}\phi\left(\frac{\bar{y}_0 - \mu_0}{\sigma_2}\right)\frac{1}{\sigma_1}\phi\left(\frac{\bar{x}_0 - \mu_0}{\sigma_1}\right)\psi(\bar{y}_0, \bar{x}_1'), \quad (5.12)$$

where $\psi(\bar{y}_1, \bar{x}_1') = \frac{1}{\sigma_2}\phi\left(\frac{\bar{y}_1 - \mu_1}{\sigma_2}\right)\prod_{k=1}^{K}\frac{1}{\sigma_1}\phi\left(\frac{\bar{x}_k - \mu_k}{\sigma_1}\right)$ and $\bar{x}_1' = (\bar{x}_1, ..., \bar{x}_K)$.

Expanding the terms in this density and using the result in expression (3.11), this density can be re-expressed as

$$\begin{aligned}
&f(\bar{y}_1, \bar{y}_0, \bar{x}|Q) \\
&= \frac{I_{[Q(x)]}}{P(\mu)}\phi\left(\frac{\frac{\sigma_2}{\sigma_1}\bar{x}_0 + \frac{\sigma_1}{\sigma_2}\bar{y}_0 - \mu_0\alpha_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)\frac{1}{\sigma_1^2}\phi\left(\frac{\bar{x}_0 - \left(\frac{\sigma_2}{\sigma_1}\bar{x}_0 + \frac{\sigma_1}{\sigma_2}\bar{y}_0\right)/\alpha_1}{\frac{\sigma_1^2}{\sigma_2^2}\alpha_2}\right)\psi(\bar{y}_1, \bar{x}_1'). \qquad (5.13)
\end{aligned}$$

Let $Z_0 = \left(\frac{\sigma_2}{\sigma_1}\bar{X}_0 + \frac{\sigma_1}{\sigma_2}\bar{Y}_0\right)$, then by the factorisation criterion (Theorem 3.4.6) it

follows that $T_0 = (\bar{X}_1, ..., \bar{X}_K, \bar{Y}_1, Z_1)$ is a sufficient and complete statistic for μ_0. Hence, the UMVCUE is found by $E[\bar{Y}_0|T_0, Q]$ and so we seek the densities $f(\bar{y}_0, \bar{x}_1, ..., \bar{x}_K, \bar{y}_1, z_0)$ and $f(\bar{x}_1, ..., \bar{x}_K, \bar{y}_1, z_0)$. In order to derive these densities, we follow similar steps to the derivation of densities (5.8) and (5.9), noting that $\bar{x}_0$ is unbounded. The required densities are

$$f(\bar{y}_0, \bar{x}_1, ..., \bar{x}_K, \bar{y}_1, z_0) = \frac{I_{[Q(x)]}}{P(\mu)} \phi \left(\frac{z_0 - \mu_0 \alpha_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right) \frac{\alpha_2}{\sigma_2^2} \psi(\bar{y}_1, \bar{x}_1') \tag{5.14}$$

and

$$f(\bar{x}_1, ..., \bar{x}_K, \bar{y}_1, z_0) = \frac{I_{[Q(x)]}}{P(\mu)} \phi \left(\frac{z_0 - \mu_0 \alpha_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right) \frac{1}{\sigma_2^2} \phi \left(\frac{\bar{y}_0 - \frac{z_0}{\alpha_1}}{\alpha_2} \right) \psi(\bar{y}_1, \bar{x}_1'). \tag{5.15}$$

Thus, following expressions (3.8) and (3.9) of a truncated normal, the UMVCUE for μ_0 is

$$\begin{aligned} \tilde{\mu}_0 = E[\bar{Y}_0|T_0, Q] &= \int_{-\infty}^{\infty} \frac{\bar{y}_0}{\alpha_2} \alpha_2 \phi \left(\frac{\bar{y}_0 - z_0/\alpha_1}{\alpha_2} \right) \mathrm{d}\bar{y}_0 \\ &= \frac{z_0}{\alpha_1} \\ &= \hat{\mu}_0. \end{aligned} \tag{5.16}$$

Hence, from expressions (5.11) and (5.16) an unbiased estimator for θ_s, accounting for treatment selection is

$$\tilde{\theta}_s = \tilde{\mu}_s - \tilde{\mu}_0. \tag{5.17}$$

As can be seen in expression (5.16), the UMVCUE for the control treatment is equal to the naive estimator $\hat{\mu}_0$. Hence, this indicates that bias in $\tilde{\theta}_s$ can only be attributed to the naive estimator for the selected experimental treatment $\hat{\mu}_s$. Furthermore, since no early stopping for futility is considered, $\tilde{\mu}_1$ is the same as the Cohen and Sackrowitz [1989] estimator.

5.7 Simulation study

To compare the properties of the naive MLE with the conditionally unbiased estimator described in the preceding section from Kimani et al. [2013], a simulation study is conducted with three scenarios. The first scenario assumes equal treatment means, the second assumes one experimental treatment is truly superior to all other experimental treatments and the third scenario assumes unequal treatment means. In each scenario, we consider a trial with 4 treatment groups and set the total numbers of patients per arm to $n = 400$ for 500,000 replications. Recall the number of patients in stage 1 is defined by $n_1 = \tau n$, where τ denotes the selection time. For each scenario, Figure 5.4 shows bias and RMSE in units of standard error of the naive MLE $\hat{\theta}_s$ and the conditionally unbiased estimator $\tilde{\theta}_s$, given by expressions (5.3) and (5.17), respectively.

For the case of equal treatment means, similar properties of the naive estimator and UMVCUE have been illustrated by Kimani et al. [2013]. However, since they are primarily interested in the influence of a futility boundary, their study explores only small effect sizes with up to 5 experimental treatment arms for a range of futility stopping boundaries. Due to the futility boundary, their results indicate a larger bias in the naive estimator compared to results shown here for the case of equal treatment means in Figures 5.1 and 5.4, as these do not consider early stopping for futility. We therefore examine properties of the estimators for not only equal treatment means, but also for unequal treatment means and a range of experimental treatment arms with small to large effect sizes, thus adding to the results in Kimani et al. [2013].

Bias plots (Figure 5.4, top row) confirm the new UMVCUE has zero bias for all selection times, in all cases. In terms of RMSE (Figure 5.4, bottom row), for the case of equal means, it can be seen that the RMSE for $\tilde{\theta}_s$ increases with selection

time with a particularly steep increase for selection time above 0.8. As $\tilde{\theta}_s$ is unbiased, its RMSE is its variance for all selection time, which is partly influenced by the bias correction term. Therefore, this increase in RMSE for later selection time is partly due to the more variable stage 2 estimate.

Now considering the case when one treatment is truly more effective than all other experimental treatments, Figure 5.4 shows that the UMVCUE RMSE approaches the MLE RMSE. This is because the correction term in the UMVCUE tends to zero as the difference between the selected and second largest mean increases. In the case of unequal means, the UMVCUE has a larger RMSE compared to the MLE with a gradual increase observed with selection time.

Hence, as similarly shown by Kimani et al. [2013] for the case of equal treatment means, the UMVCUE has favourable properties in terms of bias and RMSE for early selection times, with a steep increase in RMSE observed for late selection times. In contrast, we have shown that when all treatment means are different, the increase in RMSE for late selection times is not as steep compared with the case of equal means. Furthermore, when there is one superior mean, the RMSE of the naive and unbiased estimators are comparable for all selection times for the case of 4 experimental treatment arms. We have additionally shown that bias in the naive estimator is overall smaller when there are unequal treatment means or when only one treatment is the most effective compared to the case of all equal means.

These results highlight that depending on the selection time and mean effect size, a trade-off of bias and RMSE should be considered; as a small amount of bias in a more precise estimator may be preferable to a larger MSE in an unbiased estimator.

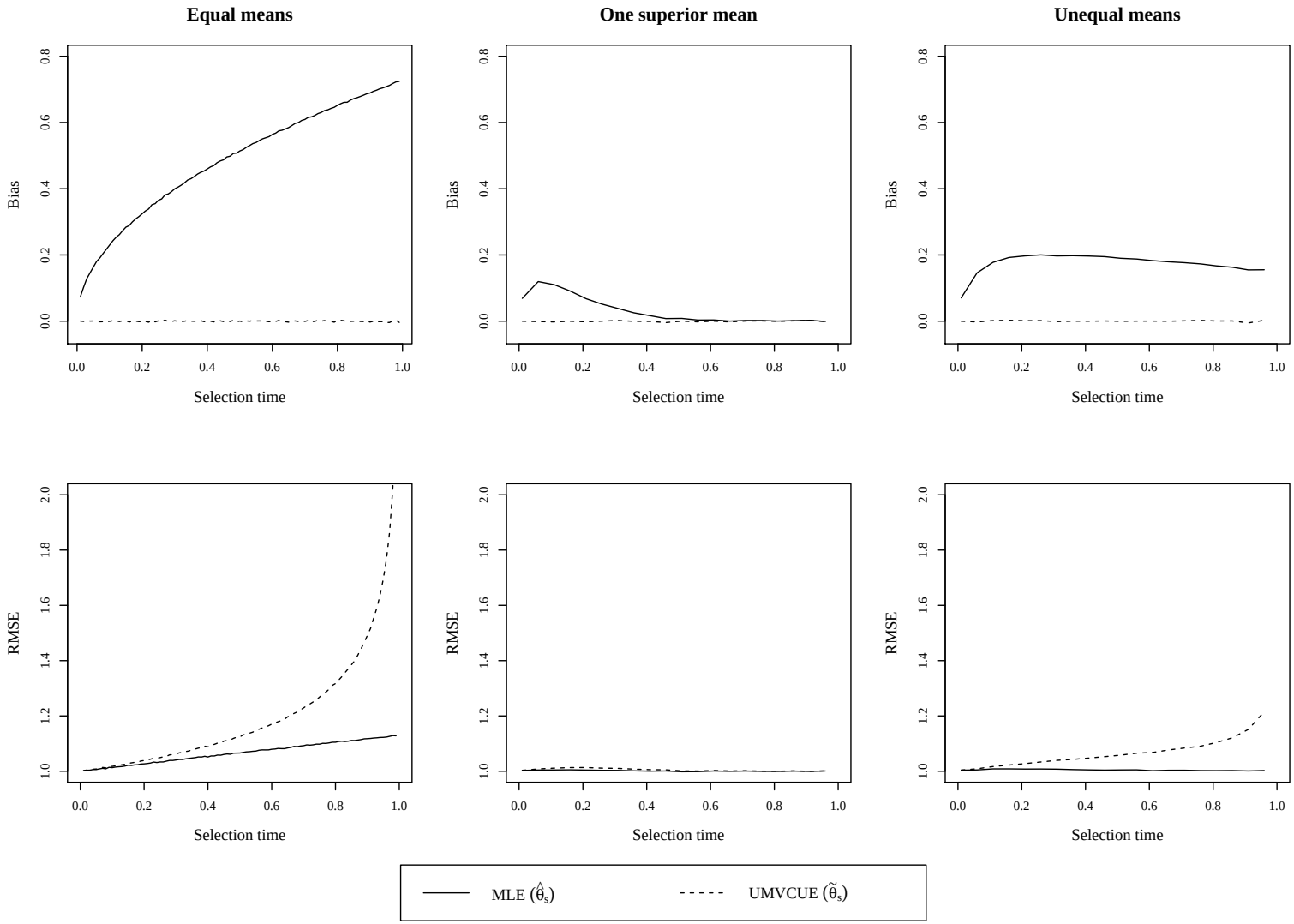

Figure 5.4: Bias (top row) and RMSE (bottom row) in units of $\sigma\sqrt{2/n}$ of $\hat{\theta}_s$ and $\tilde{\theta}_s$ for the scenario of 4 experimental treatment arms. Plots show the average of 500,000 simulations for each value of the selection time τ. Each column represents a different combination of treatment means μ_k $(k = 1, ..., 4)$: for equal means, $\mu_k = 0$; one superior mean, $\mu_1 = 0.3$ and $\mu_{k\neq 1} = 0$; and unequal means, $\mu_k = 0.1k$.

5.8 Conclusion

This chapter presented methods for unbiased estimation of normal outcomes following treatment selection in seamless phase II/III clinical trials. Properties of the naive maximum likelihood estimator were explored for various trial scenarios, including a range of effect sizes, number of experimental treatment arms and selection times. It was shown that existing methods for unbiased and efficient estimation utilise the Rao-Blackwell theorem to derive uniformly minimum variance unbiased estimators, conditional on treatment selection. A simulation study comparing an unbiased estimator with the naive estimator indicated that UMVCUEs are desirable for certain scenarios, where a trade-off of bias and RMSE should be considered.

Since these methods assume uncorrelated stage 1 and stage 2 data, the estimators described in this chapter are not appropriate for use in the analysis of time-to-event data with treatment selection. The following chapter therefore builds on the work in this chapter to address the issues with time-to-event data analysis in seamless phase II/III clinical trials.

Chapter 6

Correcting for treatment selection bias and correlation in seamless phase II/III clinical trials with time-to-event data

Despite methods of estimation described in the previous chapter correcting for selection bias in the case of normally distributed outcomes, additional challenges arise when analysis is based on time-to-event data. This chapter aims to develop unbiased estimators accounting for treatment selection for time-to-event outcomes in seamless phase II/III clinical trials.

The first section introduces the setting and describes the issues with analysis of time-to-event data in two-stage trials. This is followed by a simulation study to assess the degree of selection bias for different trial scenarios and thus motivate the need for unbiased estimators. The following section describes the concept of independent increments for use in subsequent sections where asymptotically unbiased estimators

are derived for time-to-event outcomes. The properties of the estimators developed are then presented by simulation for various trial scenarios. This chapter concludes with a summary on estimation correcting for treatment selection bias and correlation in seamless phase II/III clinical trials with time-to-event data.

6.1 Setting

As described in Section 2.4.1, seamless phase II/III clinical trials consist of two distinct stages with an interim analysis conducted part-way through the trial. In stage 1, patients are recruited to K experimental treatment arms plus a control arm. At a predefined point in time, an interim analysis is conducted to select the most efficacious experimental treatment based on a predefined selection rule.

Figure 6.1 depicts the possible outcomes for time-to-event endpoints in this two-stage setting. During stage 1, patients may experience the event of interest (black circle, e.g. patient 1) or they may be censored if they are either lost to follow-up (white circle, e.g. patient 2) or yet to experience the event of interest, in which case they are still in the trial at the time of interim analysis (e.g. grey circle, patients 3 and 4). If the latter occurs, then follow-up of these patients continues in stage 2, where they either go on to experience the event of interest (e.g. patient 3) or are lost to follow-up (e.g. patient 4). In this thesis, such cases, where follow-up continues in stage 2, are referred to as delayed events.

In stage 2, new patients are recruited to the selected experimental treatment and control group. Again, these patients may experience the event of interest before the final analysis (e.g. patient 5) or may be censored at the time last known to be event free if they are either lost to follow-up (e.g. patient 6) or do not experience the event at the time of final analysis (e.g. patient 7). At the end of the trial, interest lies in estimation of the treatment effect using data accrued in both stages of the

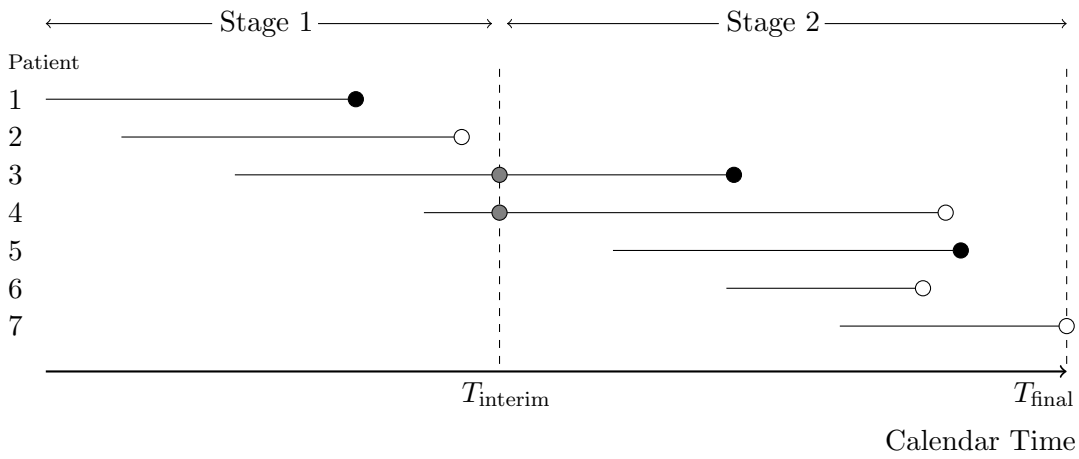

Figure 6.1: Example of survival data in a two-stage clinical trial. Black circles denote an event, and grey and white circles denote censored observations at the interim and final analysis, respectively. T_{interim} and T_{final} denote the calendar time of the interim and final analysis, respectively.

trial.

With time-to-event data, a patient's survival time is defined as the minimum of the observed censoring and event times. At the interim analysis, the stage 1 estimator is based on data accrued up until the time of interim analysis for all experimental treatment and control arms. At the final analysis, the stage 2 estimator is based on all accumulated data in the trial for the selected experimental treatment and control arms. Defining the stage 2 estimator in this way leads to correlated stage 1 and 2 data. This is due to delayed events as the stage 2 estimator consists of patients recruited in stage 1. Although defining the stage 2 estimator to be based only on new patients followed up in stage 2 would eliminate this correlation, it would also ignore the additional follow-up of stage 1 patients which would be an inefficient use of data. Hence, additional follow-up of stage 1 patients is included in the analysis but the common assumption of independence as with normally distributed data is violated. This therefore may result in a biased estimator at the end of the trial.

Furthermore, unlike the standard setting with normally distributed outcomes, treat-

ment effects are quantified by the relative difference of hazards in the experimental group compared to the control group, rather than the absolute mean difference. Hence, when analysis is based on semi-parametric approaches, such as Cox regression analysis described in Chapter 4, current methods cannot be applied directly as estimates are correlated and not normally distributed.

Due to the added complexities in the analysis of time-to-event data described here, the main limitations of the estimators discussed in the previous chapter for application to time-to-event to event data are now described. The Bowden and Glimm [2008] estimator accounts for selection but does not assume correlation within and between stages. The Kimani et al. [2013] estimator is similar to the Bowden and Glimm estimator in that it accounts for selection, ignoring within-stage and between-stage correlation, but additionally allows for early stopping. The Robertson et al. [2016] estimator accounts for selection and correlation within stage 1 but does not assume correlation between stages. In addition, all of these estimators assume data are normally distributed with treatment effect sizes ordered in increasing magnitude. Therefore, by extension of these methods, new estimators are developed in this chapter that account for treatment selection bias in the analysis of two-stage time-to-event data. Development of these estimators first focusses on extension of the Kimani et al. [2013] estimator, where separate control arms are assumed for each experimental treatment arm in Section 6.7. Then the work by Robertson et al. [2016] is considered with an estimator assuming a common control arm derived in Section 6.8.

6.2 Selection rule

Clinical trials with time-to-event endpoints are generally event driven which means they are designed with the expected number of events as opposed to number of

patients. Therefore, in contrast to the previous chapter, for time-to-event data the information fraction, τ, depends on the number of events. Let d denote the total number of events that occur at the interim analysis across all experimental treatment and control arms and D denote the total number of events at the final analysis for the entire trial. Then the information fraction is defined by $\tau = \frac{d}{D}$.

Let θ_k denote the true log HR for experimental treatment k ($k = 1, ..., K$) and $\hat{\theta}_{ik}$ denote the estimator for θ_k in stage i ($i = 1, 2$). At the interim analysis, we consider a selection rule similar to that in Section 5.3. However, as treatment efficacy is now quantified by the log HR, the most efficacious treatment is defined as that which has the smallest observed log HR. Hence at the interim analysis, treatments are ordered in decreasing magnitude of effect size such that the estimated log HR of the selected experimental treatment is denoted by $\hat{\theta}_{1s} = \min\{\hat{\theta}_{1k}\}$, $s \in \{1, ..., K\}$.

6.3 Simulation study assessing bias of the naive estimator

As we are interested in unbiased estimation, a simulation study is conducted to investigate the degree of selection bias of the naive estimator $\hat{\theta}_s$ calculated at the end of the trial. We assess the influence of the number of experimental treatments, as well as selection time for three scenarios of effect sizes. The selection time is based on the information fraction τ as defined in the preceding section. Up to $K = 6$ experimental treatments are considered with $\tau \in \{0.1, ..., 0.9\}$. A total of $n = 1500$ patients are assigned to the control arm and each experimental treatment arm k ($k = 1, ..., K$). The target number of events at the final analysis is defined as $D = 900$ such that the number of events at the interim analysis in each scenario is defined by $d = \tau D$. As described in Chapter 4, different distribution functions may be utilised that best describe survival times. Since the Weibull density function has

been shown to have the most flexible form, which allows the hazard to vary with time, we assume survival times follow a Weibull distribution.

Selection bias and RMSE of the naive estimator $\hat{\theta}_s$ are calculated from 10,000 simulated studies for each τ. Selection bias for the selected treatment s is calculated as the mean of the differences in the estimated log HR and the true log HR, that is, $\hat{\theta}_s - \theta_s$. The MSE is calculated as the mean of the squared differences between $\hat{\theta}_s$ and θ_s. Recall, since $\theta_k < 0$ $(k = 1, ..., K)$ implies superiority for treatment k, a negative bias implies a positively biased estimator, that is, $\hat{\theta}_s$ overestimates the true treatment effect.

6.3.1 Equal log HRs

Two scenarios of equal log HRs are investigated. The first considers the null case where all experimental treatments are equally ineffective, where $\theta_k = 0$ $(k = 1, ..., K)$. Event times, E_j, are simulated for each patient $j = 1, ..., n$ in control group c and treatment group k $(k = 1, ..., K)$ from a Weibull distribution with rate and shape parameters set to $\lambda_c = \lambda_k = 0.15$ and $\gamma_c = \gamma_k = 2.5$, respectively. Additionally, censoring times, C_j, are simulated for the control and treatment arms from Weibull$(\lambda_c/1.5, 1)$ and Weibull$(\lambda_k/1.5, 1)$, so that a patient's survival time at the point of analysis is defined as the minimum of the simulated event and censoring times, that is, $\min\{E_j, C_j\}$. An event is assumed to occur if $E_j < C_j$, and thus at the time of analysis, if a patients survival time is greater than the calendar time of analysis, the event is defined as censored.

The second scenario assumes all treatments are equally effective with $\theta_k = -0.22$ $\forall k$; this equates to a HR of 0.8. In this case, event and censoring times are simulated as described above with Weibull parameters $\lambda_c = 0.15$, $\lambda_k = 0.12$ and $\gamma_c = \gamma_k = 2.5$.

At the interim analysis, an estimate of the log HR, $\hat{\theta}_{1k}$, is computed from the log-rank score statistic for each treatment group at each selection time τ. Following the selection rule in Section 6.2, the treatment with the smallest estimated log HR, $\hat{\theta}_{1s}$, is deemed most efficacious and is thus selected. Survival times for additional patients are then simulated to this selected treatment arm and control arm. The final analysis is conducted after a total of $D = 900$ events are observed in the entire trial, that is, after a total of $D(1-\tau)$ new stage 2 events on the selected treatment and control arms. At the final analysis, an overall estimate of the log HR for the selected treatment is calculated from all data accrued in the trial. This is denoted by $\hat{\theta}_s$ and is referred to as the naive estimator as it does not adjust for treatment selection at the interim analysis or both the correlation between stage 1 estimates and the correlation due to censored observations at the interim analysis.

Figure 6.2(a) shows the bias (top) and RMSE (bottom) of the estimated log HR, $\hat{\theta}_s$, as a function of selection time for up to 6 treatment groups, where all treatments are assumed equally ineffective. It can be seen that as the number of treatment comparisons increases, that is as K increases, selection bias and RMSE of the log HR increases with selection time. This is analogous to selection bias in the naive estimator for normally distributed outcomes (see Figure 5.1(a)). Hence, the naive estimator of the selected treatment overestimates the true treatment effect due to selection at the interim analysis. As expected for the case of one experimental treatment, since no selection is made, the naive estimator is unbiased for all selection times (solid line).

Figure 6.2(b) shows bias and RMSE for the scenario of equally effective treatments ($\theta_k = -0.22 \forall k$). As compared to scenario (a), a similar pattern of an increase in bias as the number of experimental treatments increases is observed, with bias also increasing with selection time for $K \geq 2$. However, the magnitude of bias is overall smaller than that for the scenario of equally ineffective treatments. Additionally,

for $K = 1$, we would expect zero bias for all selection time since there is no bias attributed to treatment selection. However, as can be seen by the solid line, there is a small amount of bias for all selection times, which indicates the naive estimator underestimates the true log HR. This is the inherent bias due to the normal approximation of the log-rank score statistic used in estimation. Therefore, this bias serves as a benchmark for the bias due to the asymptotic normality assumption and thus, asymptotically unbiased estimators developed later in this chapter, which address the problem of selection bias, are expected to be only as biased as this naive estimator.

6.3.2 Unequal log HRs

Now consider the case where one experimental treatment is superior to all other experimental treatments. Data assuming Weibull survival times are simulated as described in Section 6.3.1 with parameters $\lambda_c = 0.15$, $\lambda_1 = 0.12$, $\lambda_k = 0.15$ and $\gamma_c = \gamma_1 = \gamma_k = 2.5$, for $k \in \{2, ..., 6\}$. This assumes the most effective experimental treatment has a true HR of 0.8.

Plots of bias (top) and RMSE (bottom) are presented in Figure 6.2(c). For the case where $K = 1$, a small underestimation can be seen when selection is made early in the trial. This is similar to that observed for Scenario (b), which we believe is due to the normal approximation of the log-rank score statistic. However, for $K \geq 2$, since there is a clear superior treatment, selection bias in comparison to scenarios (a) and (b) is considerably smaller for all values of K and τ, as expected. In particular, bias increases with the number of experimental treatments, however it is approximately constant for all selection times.

For all scenarios, RMSE increases with the number of experimental treatments. Additionally, a gradual increase in RMSE is observed for increasing selection time, with

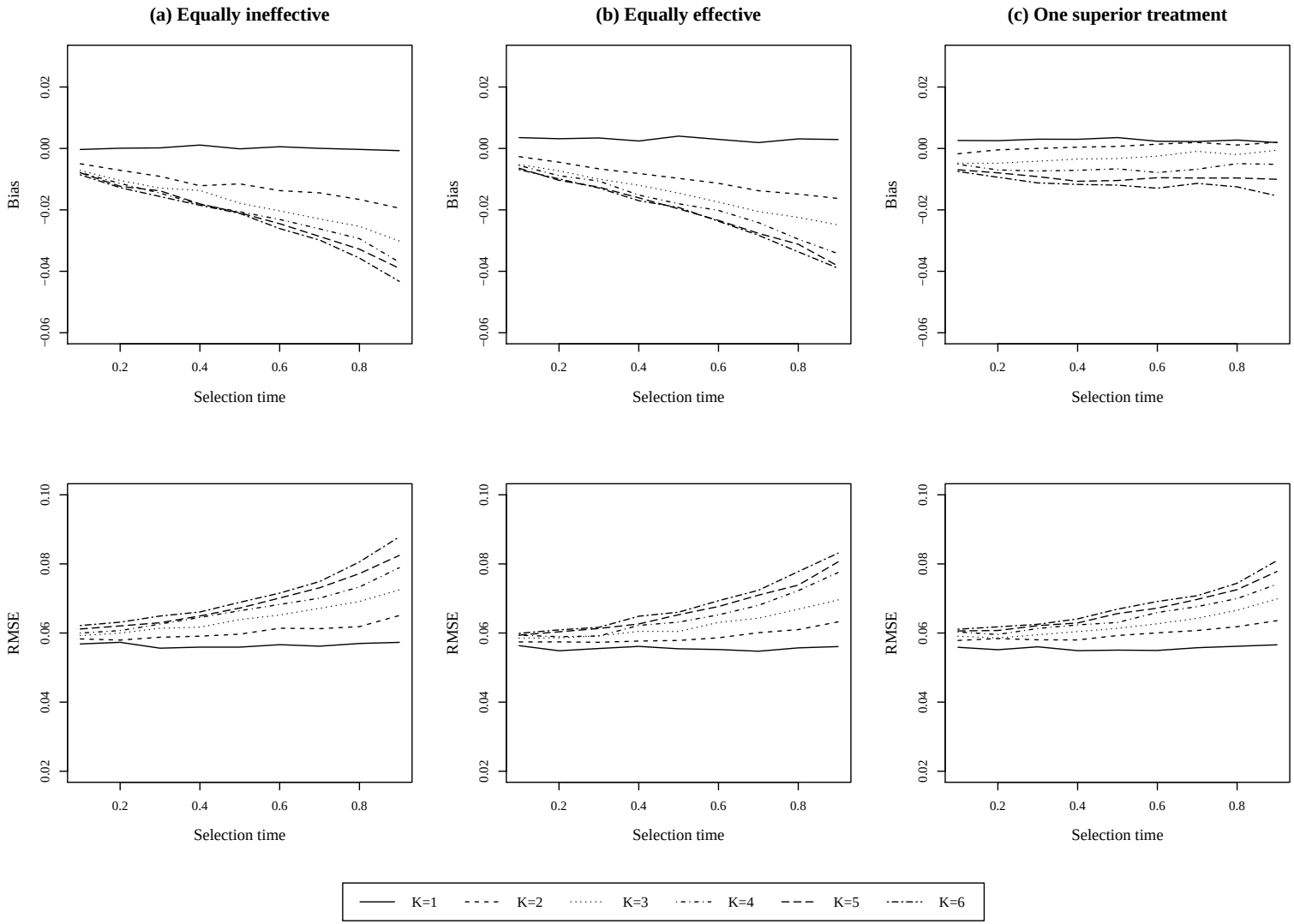

Figure 6.2: Bias (top row) and RMSE (bottom row) of the naive estimator for $K \in \{1, ..., 6\}$ treatment groups and selection time $\tau \in \{0, ..., 0.9\}$ from 10,000 simulations. Each column represents a different combination of effect sizes: (a) $\theta_k = 0 \; \forall k \in \{1, ..., K\}$, (b) $\theta_k = -0.22 \; \forall k \in \{1, ..., K\}$, and (c) $\theta_1 = -0.22$ and $\theta_k = 0 \; \forall k \in \{2, ..., K\}$.

a steeper increase observed for later selection times. This is due to the correlation between stage 1 and stage 2 data, as later selection times correspond to a larger information fraction, which means a greater proportion of events from patients recruited in stage 1 are included in the final estimate. This correlation therefore needs to be accounted for in order to obtain an efficient estimator, particularly for later selection times.

In summary, this simulation study has shown that the naive estimator is biased due to both treatment selection at the interim analysis and the correlation between stage 1 and stage 2 data. In particular, the naive estimator will overestimate the true effect size. However, in the case of one experimental treatment, the naive estimator underestimates the effect size due to the normality assumption of the log-rank test statistic. Hence, in addition to the inherent bias of the normal approximation, bias and RMSE for time-to-event outcomes depend on treatment selection at the interim analysis and correlation between stage 1 and stage 2 data.

6.4 Independent increments structure

An increment estimate, computed as the difference between the interim and final analysis estimates based on log-rank score statistics, is assumed to be independent and normally distributed [Wassmer, 2006]. Thus, defining such a estimate ensures independence of stagewise estimates, where the interim estimate is based on all data accrued up until the time of interim analysis, and the final estimate is based on all data accrued in the entire trial up until the final analysis.

Tsiatis [1982] first introduced the concept of independent increments for testing of censored survival data for the case of one experimental treatment compared to a control. He assumed the setting of one interim analysis conducted for assessment of futility or efficacy and showed the asymptotic joint distribution of the

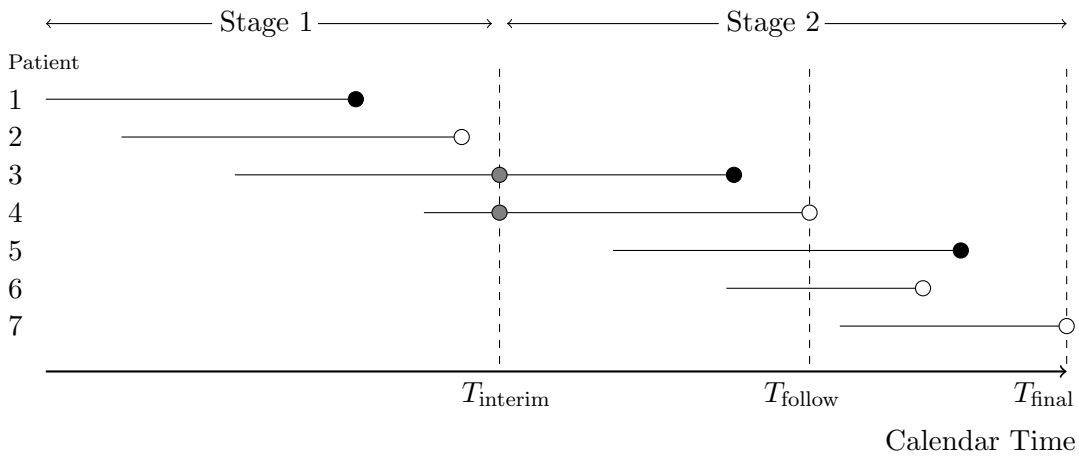

Figure 6.3: Example of survival data in a two-stage clinical trial. Black circles denote an event, and grey and white circles denote censored observations at the interim and final analysis, respectively. T_{interim} and T_{final} denote the calendar time of the interim and final analysis, respectively. T_{follow} denotes the calendar time of maximum follow-up for patients recruited in stage 1.

score test is multivariate normal with independent increments when patients are recruited randomly during the trial. Use of the independent increment structure for censored time-to-event outcomes has since been beneficial for hypothesis testing with time-to-event data. However, papers including Bauer and Posch [2004] and Jahn-Eimermacher and Ingel [2009] have highlighted the potential limitations of this assumption in adaptive trials.

More recently, methods to ensure independence of stagewise test statistics, primarily for the problem of multiple testing in two-stage survival trials have been proposed by several authors including Di Scala and Glimm [2011], Jenkins et al. [2011] and Irle and Schäfer [2012]. As this thesis focusses on treatment effect estimation using the log-rank score statistic, these methods for multiple hypothesis testing are only discussed briefly in order to highlight the concepts that will be used for point estimation in subsequent sections. A comprehensive discussion on group sequential tests and adaptive design methodology for hypothesis testing can be found in Jennison and Turnbull [2000], Wassmer and Brannath [2016] and Whitehead [1997].

The main requirement that has been highlighted in the literature is that design modifications must be independent of the primary outcome at the interim analysis, in order to avoid the possibility of predicting survival of the remaining patients at risk [Jenkins et al., 2011]. This is because information from patients with delayed events may be correlated with the survival time at the final analysis. This means that at the final analysis, the additional follow-up of patients with delayed events at the interim analysis must be prefixed in order to ensure the variance-covariance matrix of the test statistics remains unchanged [Di Scala and Glimm, 2011]. Therefore, we define T_{follow} to be the maximum follow-up time for patients recruited prior to T_{interim}, as illustrated in Figure 6.3. This implies that any events that occur after T_{follow} for patients recruited in stage 1 are effectively ignored as survival times are censored at T_{follow}, for example, Figure 6.3, patient 4. Hence, the choice of T_{follow} needs to be tailored for each trial, as it would not be desirable to pre-specify T_{follow} to allow many extra events from stage 1 patients that were censored at the interim analysis after T_{follow} in the case where all events for patients recruited in stage 2 are observed well before observing the pre-specified extra number of events from stage 1 patients. Conversely, pre-specifying few extra events from stage 1 patients, risks losing information that would be obtained from the censored observation.

6.5 Overview of available methods for analysing time-to-event data in seamless phase II/III clinical trials

There is vast literature available which addresses the multiplicity problems for time-to-event data. See, for example, Xi et al. [2016] for an overview of methods for multiplicity issues in cancer clinical trials. Since asymptotic normality of the log-rank score statistic is a well known assumption for large samples, the independent

increments structure has allowed group sequential hypothesis testing to be based on the log-rank test [Desseaux and Porcher, 2007, Jahn-Eimermacher and Ingel, 2009, Schäfer and Müller, 2001, Wassmer, 2006].

As decribed in the previous section, we utilise the independent increments structure with a prefixed follow-up time for delayed events. For hypothesis testing based on the standardised log-rank test statistic, Di Scala and Glimm [2011] give the joint distribution of a group sequential log-rank test for a two-stage two-arm trial with correlated time-to-event outcomes. They use the independent increments assumption to obtain an independent increments covariance structure for the joint distribution of stage 1 and 2 statistics. Although the use of independent increments has only been considered for hypothesis testing in group sequential designs, their use may be extended to point estimation based on the log-rank score statistic. Therefore, the joint structure of the correlated test statistics is now described, as this concept will be used for the derivation of point estimators in Sections 6.7 and 6.8.

6.5.1 Joint distribution of log-rank test statistics

As before, let θ_k denote the log HR for treatment k $(k = 1, ..., K)$ compared to control. The score statistic given in Section 4.3.1 may be used to obtain an estimate of the log HR. This is the sum of the difference of the observed and expected events summed over all event times in the trial. As we are considering a two-stage trial, at the interim and final analyses, a log-rank statistic is now computed for each stage based on all accumulated data at the end of each stage i $(i = 1, 2)$. For now, we assume all treatment arms are continued into the second stage.

Data for each experimental treatment and control arm at the j^{th} death time are summarised in Table 6.1. For distinct event times r_i in stage i, the number of deaths that occur in group k at time t_j $(j = 1, ..., r_i)$ is denoted by d_{jk}. It is known

Table 6.1: Data at the j^{th} death time for up to K treatment groups

Treatment group	**Observed no. of deaths**	**No. alive**	**No. at risk**
Control	d_{j0}	$n_{j0} - d_{j0}$	n_{j0}
1	d_{j1}	$n_{j1} - d_{j1}$	n_{j1}
2	d_{j2}	$n_{j2} - d_{j2}$	n_{j2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
K	d_{jK}	$n_{jK} - d_{jK}$	n_{jK}

that the conditional distribution of d_{jk} given r_1, r_2 and $(d_{j0}+d_{jk})$ is hypergeometric with expected number of deaths in group k at time t_j given by

$$e_{jk} = \frac{n_{jk}(d_{j0} + d_{jk})}{n_{j0} + n_{jk}}.$$

Hence, the score statistic in stage i for experimental treatment k compared to control is given by

$$S_{ik} = \sum_{j=1}^{r_i} (d_{jk} - e_{jk}). \tag{6.1}$$

As mentioned in Section 4.3.1, the asymptotic distribution of S_{ik} is normal with mean $\theta_k V_{ik}$ and variance V_{ik}, which is the sum of the variance of d_{jk} over all distinct event times for all data accumulated in stage i. This is given by

$$V_{ik} = \sum_{j=1}^{r_i} p_{jk}(1 - p_{jk}), \tag{6.2}$$

where, as described in Section 4.3.1, p_{jk} is defined by

$$p_{jk} = \begin{cases} \frac{n_{jk}}{n_{j0}+n_{jk}}, & \text{if event occured in the control or treatment group } k \text{ at time } t_j, \\ 0, & \text{otherwise.} \end{cases} \tag{6.3}$$

The standardised log-rank test statistic for treatment k in stage i is then given by

$$Z_{ik} = \frac{S_{ik}}{\sqrt{V_{ik}}} \sim N(\theta_k \sqrt{V_{ik}}, 1).$$

This statistic is most commonly used for group sequential hypothesis testing since it is standard normal under the null [Di Scala and Glimm, 2011, Irle and Schäfer, 2012, Jenkins et al., 2011]. For large samples the variance is approximately a quarter of the total number of observed events, that is, $V_{ik} = \frac{\sum_{j=1}^{r_i}(d_{j0}+d_{jk})}{4}$.

At the interim and final analyses, each experimental treatment is compared to the common control arm, which means estimates at each analysis are correlated. This is referred to as *within-stage correlation*. Let $V_{1,k,l}$ denote the covariance of stage 1 statistics Z_{1k} and Z_{1l} for treatments $k, l \in \{1, ..., K\}, k \neq l$. Then from Di Scala and Glimm [2011],

$$V_{1,k,l} = \frac{\sum_{j=1}^{r_1} \phi_j}{\sqrt{\sum_{j=1}^{r_1} p_{jk}(1-p_{jk}) \cdot \sum_{j=1}^{r_1} p_{jl}(1-p_{jl})}}, \tag{6.4}$$

where $\phi_j = p_{jk}p_{jl}$ is the correlation of the score statistics.

Additionally, as mentioned, delayed events from stage 1 patients cause stagewise statistics, Z_{1k} and Z_{2k}, to be correlated. This correlation is denoted by ρ_k and is obtained as follows. It is known that, $cov(S_{1k}, S_{2k}) = V_{1k}$ and let $\psi_{1,k,l}$ denote $cov(S_{1k}, S_{1l}) = \sum_{j=1}^{r_1} \phi_j$. Then

$$\begin{aligned} \rho_k = cov(Z_{1k}, Z_{2k}) &= cov\left(\frac{S_{1k}}{\sqrt{V_{1k}}}, \frac{S_{2k}}{\sqrt{V_{2k}}}\right) \\ &= \frac{1}{\sqrt{V_{1k}V_{2k}}} cov(S_{1k}, S_{2k}) \\ &= \frac{V_{1k}}{\sqrt{V_{1k}V_{2k}}} \\ &= \sqrt{\frac{V_{1k}}{V_{2k}}} \end{aligned} \tag{6.5}$$

and

$$\begin{aligned} cov(Z_{1k}, Z_{2l}) &= cov\left(\frac{S_{1k}}{\sqrt{V_{1k}}}, \frac{S_{2l}}{\sqrt{V_{2l}}}\right) \\ &= \frac{\psi_{1,k,l}}{\sqrt{V_{1k}V_{2l}}} \\ &= \frac{\psi_{1,k,l}}{\sqrt{V_{1k}V_{1l}}}\sqrt{\frac{V_{1l}}{V_{2l}}} \\ &= \rho_l V_{1,k,l}. \end{aligned} \tag{6.6}$$

It follows that the asymptotic joint distribution of stage 1 and 2 statistics can be written as

$$\begin{pmatrix} Z_{11} \\ Z_{12} \\ Z_{13} \\ \vdots \\ Z_{1K} \\ Z_{21} \\ Z_{22} \\ Z_{23} \\ \vdots \\ Z_{2K} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_1\sqrt{V_{11}} \\ \theta_2\sqrt{V_{12}} \\ \theta_3\sqrt{V_{13}} \\ \vdots \\ \theta_K\sqrt{V_{1K}} \\ \theta_1\sqrt{V_{21}} \\ \theta_2\sqrt{V_{22}} \\ \theta_3\sqrt{V_{23}} \\ \vdots \\ \theta_K\sqrt{V_{2K}} \end{pmatrix}, \begin{pmatrix} 1 & V_{1,1,2} & V_{1,1,3} & \cdots & V_{1,1,K} & \rho_1 & \rho_2 V_{1,1,2} & \rho_3 V_{1,1,3} & \cdots & \rho_K V_{1,1,K} \\ V_{1,1,2} & 1 & V_{1,2,3} & \cdots & V_{1,2,K} & \rho_2 V_{1,1,2} & \rho_2 & \rho_2 V_{1,2,3} & \cdots & \rho_K V_{1,2,K} \\ V_{1,1,3} & V_{1,2,3} & 1 & \cdots & V_{1,3,K} & \rho_3 V_{1,1,3} & \rho_2 V_{1,2,3} & \rho_3 & \cdots & \rho_K V_{1,3,K} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{1,1,K} & V_{1,2,K} & V_{1,3,K} & \cdots & 1 & \rho_K V_{1,1,K} & \rho_K V_{1,2,K} & \rho_K V_{1,3,K} & \cdots & \rho_K \\ \rho_1 & \rho_2 V_{1,1,2} & \rho_3 V_{1,1,3} & \cdots & \rho_K V_{1,1,K} & 1 & V_{2,1,2} & V_{2,1,3} & \cdots & V_{2,1,K} \\ \rho_2 V_{1,1,2} & \rho_2 & \rho_2 V_{1,2,3} & \cdots & \rho_K V_{1,2,K} & V_{2,1,2} & 1 & V_{2,2,3} & \cdots & V_{2,2,K} \\ \rho_3 V_{1,1,3} & \rho_2 V_{1,2,3} & \rho_3 & \cdots & \rho_K V_{1,3,K} & V_{2,1,3} & V_{2,2,3} & 1 & \cdots & V_{2,3,K} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_K V_{1,1,K} & \rho_K V_{1,2,K} & \rho_K V_{1,3,K} & \cdots & \rho_K & V_{2,1,K} & V_{2,2,K} & V_{2,3,K} & \cdots & 1 \end{pmatrix} \right) \tag{6.7}$$

6.5.2 Stage 2 increment statistics

Now in order to apply group sequential methods available for normally distributed data, stage 1 and stage 2 statistics need to be uncorrelated. As described previously, a well known approach for eliminating this between-stage correlation is to utilise the independent increments assumption. This states that, by linear transformation, a new statistic denoted by $\tilde{Z}_{2k}$ is approximately normal with mean $\theta_k\sqrt{V_{2k}-V_{1k}}$ and variance 1 such that $\text{cov}(Z_{1k}, \tilde{Z}_{2k}) = 0$ [Di Scala and Glimm, 2011]. This is derived as follows.

The increment score statistic is defined as $\tilde{S}_{2k} = S_{2k} - S_{1k}$ which is normally distributed with mean θ_k and variance $V_{2k} - V_{1k}$. Hence

$$\begin{aligned} \tilde{Z}_{2k} &= \frac{\tilde{S}_{2k}}{\sqrt{V_{2k}-V_{1k}}} \\ &= \frac{Z_{2k}\sqrt{V_{2k}} - Z_{1k}\sqrt{V_{1k}}}{\sqrt{V_{2k}-V_{1k}}} \end{aligned} \tag{6.8}$$

so $\text{E}(\tilde{Z}_{2k}) = \theta_k\sqrt{V_{2k}-V_{1k}}$ and $\text{var}(\tilde{Z}_{2k}) = 1$ and hence $\tilde{Z}_{2k} \sim N(\theta_k\sqrt{V_{2k}-V_{1k}}, 1)$. It follows, by linear transformation of the multivariate normal distribution (6.7), that the joint asymptotic distribution of stage 1 statistics and stage 2 increment statistics is

$$\begin{pmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1K} \\ \tilde{Z}_{21} \\ \tilde{Z}_{22} \\ \vdots \\ \tilde{Z}_{2K} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1\sqrt{V_{11}} \\ \theta_2\sqrt{V_{12}} \\ \vdots \\ \theta_K\sqrt{V_{1K}} \\ \theta_1\sqrt{V_{21}-V_{11}} \\ \theta_2\sqrt{V_{22}-V_{12}} \\ \vdots \\ \theta_K\sqrt{V_{2K}-V_{1K}} \end{pmatrix}, \begin{pmatrix} 1 & V_{1,1,2} & \cdots & V_{1,1,K} & 0 & 0 & \cdots & 0 \\ V_{1,1,2} & 1 & \cdots & V_{1,2,K} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{1,1,K} & V_{1,2,K} & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & V^*_{2,1,2} & \cdots & V^*_{2,1,K} \\ 0 & 0 & \cdots & 0 & V^*_{2,1,2} & 1 & \cdots & V^*_{2,2,K} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & V^*_{2,1,K} & V^*_{2,2,K} & \cdots & 1 \end{pmatrix} \right)$$

where the covariance of $(\tilde{Z}_{2k}, \tilde{Z}_{2l})$ is given by

$$V^*_{2,k,l} = \frac{\sum_{j=1}^{r_2} \phi_j - \sum_{j=1}^{r_1} \phi_j}{\sqrt{V_{2k} - V_{1k}}\sqrt{V_{2l} - V_{1l}}}. \tag{6.9}$$

The independent increments structure for the standardised statistics described in this section has allowed group sequential hypothesis testing to be based on the log-rank test [Schäfer and Müller, 2001, Wassmer, 2006]. However, there are a lack of methods focusing on estimation following two-stage trials with survival data. The following sections show how the methods described above can be adapted and incorporated for developing point estimators based on the log-rank test statistic for two-stage survival trials.

6.6 Derivation of the joint density of stagewise log HRs

Since we are interested in estimating the log HR, θ_k $(k = 1, ..., K)$, we re-write stage 1 statistics, Z_{1k}, and stage 2 increments, $\tilde{Z}_{2k}$, by dividing them by their respective variances $\sqrt{V_{1k}}$ and $\sqrt{V_{2k} - V_{1k}}$. Let $\hat{\theta}_{1k}$ denote the new stage 1 statistic and $\tilde{\theta}_{2k}$ denote the new stage 2 increment statistic. Then $\hat{\theta}_{1k} = \frac{Z_{1k}}{\sqrt{V_{1k}}} = \frac{S_{1k}}{V_{1k}}$ is asymptotically normal with mean θ_k and variance $\sigma^2_{1k} = \frac{1}{V_{1k}}$. Similarly,

$$\tilde{\theta}_{2k} = \frac{S_{2k} - S_{1k}}{V_{2k} - V_{1k}} \tag{6.10}$$

is asymptotically normal with mean θ_k and variance $\sigma^2_{2k} = \frac{1}{V_{2k} - V_{1k}}$.

Let $\sigma_{1,k,l}$ and $\sigma_{2,k,l}$ denote the covariance of $(\hat{\theta}_{1k}, \hat{\theta}_{1l})$ and $(\tilde{\theta}_{2k}, \tilde{\theta}_{2l})$, respectively.

Then from expressions (6.4) and (6.9), it follows that

$$\begin{aligned}\sigma_{1,k,l} = cov(\hat{\theta}_{1k}, \hat{\theta}_{1l}) &= cov\left(\frac{Z_{1k}}{\sqrt{V_{1k}}}, \frac{Z_{1l}}{\sqrt{V_{1l}}}\right) \\ &= \frac{V_{1,k,l}}{\sqrt{V_{1k}V_{1l}}}\end{aligned}$$

and

$$\begin{aligned}\sigma_{2,k,l} = cov(\tilde{\theta}_{2k}, \tilde{\theta}_{2l}) &= cov\left(\frac{\tilde{Z}_{2k}}{\sqrt{V_{2k} - V_{1k}}}, \frac{\tilde{Z}_{2l}}{\sqrt{V_{2l} - V_{1l}}}\right) \\ &= \frac{V^*_{2,k,l}}{\sqrt{(V_{2k} - V_{1k})(V_{2l} - V_{1l})}}.\end{aligned}$$

The correlation between stage 1 statistics is therefore

$$\varphi = corr(\hat{\theta}_{1k}, \hat{\theta}_{1l}) = \frac{cov(\hat{\theta}_{1k}, \hat{\theta}_{1l})}{var(\hat{\theta}_{1k}) \cdot var(\hat{\theta}_{1l})} = \frac{V_{1,k,l}}{\sqrt{V_{1k}V_{1l}}} \cdot \sqrt{V_{1k}V_{1l}} = V_{1,k,l}. \quad (6.11)$$

Hence, the aysmptotic joint density of $(\hat{\theta}_{11}, \hat{\theta}_{12}, ..., \hat{\theta}_{1K}, \tilde{\theta}_{21}, \tilde{\theta}_{22}, ..., \tilde{\theta}_{2K})'$ is

$$\begin{pmatrix} \hat{\theta}_{11} \\ \hat{\theta}_{12} \\ \vdots \\ \hat{\theta}_{1K} \\ \tilde{\theta}_{21} \\ \tilde{\theta}_{22} \\ \vdots \\ \tilde{\theta}_{2K} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \end{pmatrix}, \begin{pmatrix} \sigma^2_{11} & \sigma_{1,1,2} & \cdots & \sigma_{1,1,K} & 0 & 0 & \cdots & 0 \\ \sigma_{1,1,2} & \sigma^2_{12} & \cdots & \sigma_{1,2,K} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,1,K} & \sigma_{1,2,K} & \cdots & \sigma^2_{1K} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \sigma^2_{21} & \sigma_{2,1,2} & \cdots & \sigma_{2,1,K} \\ 0 & 0 & \cdots & 0 & \sigma_{2,1,2} & \sigma^2_{22} & \cdots & \sigma_{2,2,K} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \sigma_{2,1,K} & \sigma_{2,2,K} & \cdots & \sigma^2_{2K} \end{pmatrix}\right).$$

Now assuming the trial always continues with treatment 1, and thus equivalently dropping treatments $k = 2, ..., K$, the density for the stage 1 statistics and the stage 2 increment statistic for treatment 1 is

$$\begin{pmatrix} \hat{\theta}_{11} \\ \hat{\theta}_{12} \\ \vdots \\ \hat{\theta}_{1K} \\ \tilde{\theta}_{21} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \\ \theta_1 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^2 & \sigma_{1,1,2} & \cdots & \sigma_{1,1,K} & 0 \\ \sigma_{1,1,2} & \sigma_{12}^2 & \cdots & \sigma_{1,2,K} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{1,1,K} & \sigma_{1,2,K} & \cdots & \sigma_{1K}^2 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_{21}^2 \end{pmatrix} \right). \tag{6.12}$$

This density can now be extended for developing point estimators based on the log-rank test statistic for two-stage survival trials that account for treatment selection and correlation. For treatment selection, without loss of generality of selecting treatment 1, since the test statistics should not change with treatment selection [Di Scala and Glimm, 2011], we must fix both the number of events from new patients recruited in stage 2 and the number of delayed events from patients recruited in stage 1 but who continue follow-up in stage 2, for the selected treatment and control arms. In addition, the total number of stage 1 events across all arms must also be predefined. Furthermore, since treatment 1 is selected as the apparently most effective treatment at the interim analysis, the expected value of the stage 2 increment, $\tilde{\theta}_{21}$, is not smaller than θ_1, since at the final analysis the estimated treatment effect, including delayed events, would be expected to underestimate the true treatment effect due to the overestimation at the interim analysis. This means that result (6.12) is unchanged irrespective of the selection, except for the index of the stage 2 log HR, and thus satisfies the independent increments structure.

6.7 Bias correction with separate control arms

To compare a similar setting for time-to-event data to normally distributed data, a uniformly minimum variance conditionally unbiased estimator (UMVCUE) is de-

rived in this section following the estimator described in Section 5.6.1 from Kimani et al. [2013]. The assumptions of their estimator are that data are normally distributed, which means stagewise estimates are independent, and that the treatment selected is that with maximum observed efficacy. Additionally, since data are normally distributed, treatment means are calculated for the control group and each experimental group separately, which means stage 1 estimates are independent. Like them, in this section, we consider the case of uncorrelated stage 1 estimates by assuming a separate control arm for each experimental treatment arm, but unlike them, we allow for stage 1 and 2 estimates to be correlated by continuing follow-up of censored patients at the interim analysis.

An asymptotically UMVCUE is now derived that is appropriate for time-to-event data that directly estimates the log HR for the selected treatment. In order to correct for the correlation between stagewise statistics, the UMVCUE utilises the independent increments assumption where the stage 2 variance, σ_{21}^2, is the variance associated with the stage 2 increment estimate, as described in Section 6.6. Additionally, as mentioned, in order to maintain independence between stage 1 estimates, in this section, separate control groups are assumed and this assumption is explored further in the next section.

Consider a trial with K experimental treatment arms and a separate control arm for each experimental treatment arm. As per the selection rule defined in Section 6.2, let Q_s denote the event that $\hat{\theta}_{1(1)} < \hat{\theta}_{1(2)} < ... < \hat{\theta}_{1(K)}$. Without loss of generality assume $\hat{\theta}_{1(k)} = \hat{\theta}_{1k}$ so that treatment 1 is selected and the index of the selected experimental treatment is $s = 1$.

Since separate control arms are assumed, stage 1 estimates are now uncorrelated

which gives the density

$$\begin{pmatrix} \hat{\theta}_{11} \\ \hat{\theta}_{12} \\ \vdots \\ \hat{\theta}_{1K} \\ \tilde{\theta}_{21} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \\ \theta_1 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^2 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_{12}^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_{1K}^2 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_{21}^2 \end{pmatrix} \right). \tag{6.13}$$

In Section 6.6, independent increment statistics for the log HR were defined for stage 2. For the selected treatment in this case, the independent increment statistic is given by $\tilde{\theta}_{21}$ (6.10). This statistic corrects for the correlation due to censoring and is an unbiased but inefficient estimator for θ_1. Hence, sufficient and complete statistics are now sought in order to Rao-Blackwellise $\tilde{\theta}_{21}$ and derive an asymptotically UMVCUE for θ_1.

The joint density in (6.13) gives the distribution of stage 1 statistics and the stage 2 increment for treatment 1. Thus from this density, conditional on selection of treatment 1, the joint density is

$$f(\hat{\theta}_{11}, ..., \hat{\theta}_{1K}, \tilde{\theta}_{21} | Q_s) = \frac{I_{Q_s}(\theta)}{P(\theta)} \frac{1}{\sigma_{21}} \phi \left(\frac{\tilde{\theta}_{21} - \theta_1}{\sigma_{21}} \right) \prod_{k=1}^{K} \frac{1}{\sigma_{1k}} \phi \left(\frac{\hat{\theta}_{1k} - \theta_k}{\sigma_{1k}} \right) \tag{6.14}$$

where, $I_{Q_s}(\theta)$ is the indicator function for Q_s and $P(\theta) = Prob(I_{Q_s}(\theta) = 1)$.

To find sufficient and complete statistics for θ_1, the result in expression (3.11) (Section 3.3.5) is used to re-express the terms $\frac{1}{\sigma_{21}} \phi \left(\frac{\tilde{\theta}_{21} - \theta_1}{\sigma_{21}} \right) \frac{1}{\sigma_{11}} \phi \left(\frac{\hat{\theta}_{11} - \theta_1}{\sigma_{11}} \right)$, and thus re-write density (6.14) as

$$f(\hat{\theta}_{11}, ..., \hat{\theta}_{1K}, \tilde{\theta}_{21} | Q_s) = \frac{I_{Q_s}(\theta)}{P(\theta)} \phi \left(\frac{\alpha_1 \theta_1 - (\frac{\sigma_{21}}{\sigma_{11}} \hat{\theta}_{11} + \frac{\sigma_{11}}{\sigma_{21}} \tilde{\theta}_{21})}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}} \right) \prod_{k=2}^{K} \frac{1}{\sigma_{1k}} \phi \left(\frac{\hat{\theta}_{1k} - \theta_k}{\sigma_{1k}} \right)$$
$$\frac{1}{\sigma_{11}^2} \phi \left(\frac{\hat{\theta}_{11} - (\frac{\sigma_{21}}{\sigma_{11}} \hat{\theta}_{11} + \frac{\sigma_{11}}{\sigma_{21}} \tilde{\theta}_{21}) / \alpha_1}{\frac{\sigma_{11}^2}{\sigma_{21}^2} \alpha_2} \right),$$

where $\alpha_1 = \dfrac{\sigma_{11}}{\sigma_{21}} + \dfrac{\sigma_{21}}{\sigma_{11}}$ and $\alpha_2 = \dfrac{\sigma_{21}^2}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}}$.

Let $\tilde{\theta}_1^* = \frac{\sigma_{21}}{\sigma_{11}}\hat{\theta}_{11} + \frac{\sigma_{11}}{\sigma_{21}}\tilde{\theta}_{21}$ and $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1^*, \hat{\theta}_{12}, ..., \hat{\theta}_{1K})$. Then $f(\hat{\theta}_{11}, ..., \hat{\theta}_{1K}, \tilde{\theta}_{21}|Q_s)$ can be transformed into $f(\hat{\theta}_{11}, \tilde{\boldsymbol{\theta}}|Q_s)$, and since the Jacobian of the transformation is 1, it is given by

$$f(\hat{\theta}_{11}, \tilde{\boldsymbol{\theta}}|Q_s) = \frac{I_{Q_s}(\theta)}{P(\theta)} \phi\left(\frac{\alpha_1\theta_1 - \tilde{\theta}_1^*}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}}\right) \frac{1}{\sigma_{11}^2} \phi\left(\frac{\hat{\theta}_{11} - \tilde{\theta}_1^*/\alpha_1}{\frac{\sigma_{11}^2}{\sigma_{21}^2}\alpha_2}\right) \prod_{k=2}^{K} \frac{1}{\sigma_{1k}} \phi\left(\frac{\hat{\theta}_{1k} - \theta_k}{\sigma_{1k}}\right).$$

Hence, by the Factorisation Theorem 3.4.6 and Definition 3.4.8, from $f(\hat{\theta}_{11}, \tilde{\boldsymbol{\theta}}|Q_s)$, it follows that $\tilde{\boldsymbol{\theta}}$ is a complete, sufficient statistic for $(\theta_1, ..., \theta_K)$.

Now by the method of Rao-Blackwellisation, the UMVCUE is found by $E[\tilde{\theta}_{21}|\tilde{\boldsymbol{\theta}}, Q_s] = \int \tilde{\theta}_{21} f(\tilde{\theta}_{21}|\tilde{\boldsymbol{\theta}}, Q_s) \mathrm{d}\tilde{\theta}_{21}$. Since,

$$f(\tilde{\theta}_{21}|\tilde{\boldsymbol{\theta}}, Q_s) = \frac{f(\tilde{\theta}_{21}, \tilde{\boldsymbol{\theta}}|Q_s)}{f(\tilde{\boldsymbol{\theta}}|Q_s)}, \tag{6.15}$$

the densities required are, $f(\tilde{\theta}_{21}, \tilde{\boldsymbol{\theta}}|Q_s)$ and $f(\tilde{\boldsymbol{\theta}}|Q_s)$.

The density in the denominator is found by integrating density $f(\hat{\theta}_{11}, \tilde{\boldsymbol{\theta}}|Q_s)$ with respect to $\hat{\theta}_{11}$ as follows. By the condition of selection, Q_s, the range of $\hat{\theta}_{11}$ is $(-\infty, \hat{\theta}_{12})$, so that

$$\begin{aligned}
& f(\tilde{\boldsymbol{\theta}}|Q_s) \\
&= \int_{-\infty}^{\hat{\theta}_{12}} f(\hat{\theta}_{11}, \tilde{\boldsymbol{\theta}}|Q_s) d\hat{\theta}_{11} \\
&= \frac{I_{Q_s}(\theta)}{P(\theta)} \phi\left(\frac{\alpha_1\theta_1 - \theta_1^*}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}}\right) \frac{1}{\sigma_{11}^2} \prod_{k=2}^{K} \frac{1}{\sigma_{1k}} \phi\left(\frac{\hat{\theta}_{1k} - \theta_k}{\sigma_{1k}}\right) \int_{-\infty}^{\hat{\theta}_{12}} \phi\left(\frac{\hat{\theta}_{11} - \tilde{\theta}_1^*/\alpha_1}{\frac{\sigma_{11}^2}{\sigma_{21}^2}\alpha_2}\right) \mathrm{d}\hat{\theta}_{11} \\
&= \frac{I_{Q_s}(\theta)}{P(\theta)} \phi\left(\frac{\alpha_1\theta_1 - \theta_1^*}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}}\right) \frac{1}{\sigma_{11}^2} \prod_{k=2}^{K} \frac{1}{\sigma_{1k}} \phi\left(\frac{\hat{\theta}_{1k} - \theta_k}{\sigma_{1k}}\right) \frac{\sigma_{11}^2}{\sigma_{21}^2}\alpha_2 \Phi\left(\frac{\hat{\theta}_{12} - \tilde{\theta}_1^*/\alpha_1}{\frac{\sigma_{11}^2}{\sigma_{21}^2}\alpha_2}\right).
\end{aligned}$$

Let

$$\omega_a = \left(\frac{\hat{\theta}_{12} - \tilde{\theta}_1^*/\alpha_1}{\frac{\sigma_{11}^2}{\sigma_{21}^2}\alpha_2} \right) = \left(\hat{\theta}_{12} - \frac{\sigma_{21}^2\hat{\theta}_{11} + \sigma_{11}^2\tilde{\theta}_{21}}{\sigma_{11}^2 + \sigma_{21}^2} \right) \frac{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}}{\sigma_{11}^2},$$

then

$$f(\tilde{\boldsymbol{\theta}}|Q_s) = \frac{I_{Q_s}(\theta)}{P(\theta)} \phi \left(\frac{\alpha_1\theta_1 - \theta_1^*}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}} \right) \prod_{k=2}^{K} \frac{1}{\sigma_{1k}} \phi \left(\frac{\hat{\theta}_{1k} - \theta_k}{\sigma_{1k}} \right) \frac{\alpha_2}{\sigma_{21}^2} \Phi(\omega_a). \tag{6.16}$$

Using similar steps and following the result in Section 3.3.5, density (6.14) can be re-expressed as

$$f(\hat{\theta}_{11}, ..., \hat{\theta}_{1K}, \tilde{\theta}_{21}|Q_s) = \frac{I_{Q_s}(\theta)}{P(\theta)} \phi \left(\frac{(\frac{\sigma_{21}}{\sigma_{11}}\hat{\theta}_{11} + \frac{\sigma_{11}}{\sigma_{21}}\tilde{\theta}_{21}) - \alpha_1\theta_1}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}} \right) \prod_{k=2}^{K} \frac{1}{\sigma_{1k}} \phi \left(\frac{\hat{\theta}_{1k} - \theta_k}{\sigma_{1k}} \right)$$
$$\frac{1}{\sigma_{11}\sigma_{21}} \phi \left(\frac{\tilde{\theta}_{21} - (\frac{\sigma_{21}}{\sigma_{11}}\hat{\theta}_{11} + \frac{\sigma_{11}}{\sigma_{21}}\tilde{\theta}_{21})/\alpha_1}{\alpha_2} \right).$$

As $\tilde{\theta}_1^* = \frac{\sigma_{21}}{\sigma_{11}}\hat{\theta}_{11} + \frac{\sigma_{11}}{\sigma_{21}}\tilde{\theta}_{21}$, by transformation of random variables (Section 3.1.1), the density $f(\tilde{\theta}_{21}, \tilde{\boldsymbol{\theta}}|Q_s)$ can be derived from $f(\hat{\theta}_{11}, ..., \hat{\theta}_{1K}, \tilde{\theta}_{21}|Q_s)$ with Jacobian of the transformation being $\frac{\sigma_{11}}{\sigma_{21}}$. This gives,

$f(\tilde{\theta}_{21}, \tilde{\boldsymbol{\theta}}|Q_s)$

$$\begin{aligned} &= \frac{I_{Q_s}(\theta)}{P(\theta)} \phi \left(\frac{\tilde{\theta}_1^* - \alpha_1\theta_1}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}} \right) \frac{1}{\sigma_{11}\sigma_{21}} \phi \left(\frac{\tilde{\theta}_{21} - \tilde{\theta}_1^*/\alpha_1}{\alpha_2} \right) \prod_{k=2}^{K} \frac{1}{\sigma_{1k}} \phi \left(\frac{\hat{\theta}_{1k} - \theta_k}{\sigma_{1k}} \right) \frac{\sigma_{11}}{\sigma_{21}} \\ &= \frac{I_{Q_s}(\theta)}{P(\theta)} \phi \left(\frac{\tilde{\theta}_1^* - \alpha_1\theta_1}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}} \right) \frac{1}{\sigma_{21}^2} \phi \left(\frac{\tilde{\theta}_{21} - \tilde{\theta}_1^*/\alpha_1}{\alpha_2} \right) \prod_{k=2}^{K} \frac{1}{\sigma_{1k}} \phi \left(\frac{\hat{\theta}_{1k} - \theta_k}{\sigma_{1k}} \right). \end{aligned} \tag{6.17}$$

Hence, from expressions (6.15), (6.16) and (6.17) the main density required for the

UMVCUE is

$$\begin{aligned} f(\tilde{\theta}_{21}|\tilde{\boldsymbol{\theta}}, Q_s) &= \frac{f(\tilde{\theta}_{21}, \tilde{\boldsymbol{\theta}}|Q_s)}{f(\tilde{\boldsymbol{\theta}}|Q_s)} \\ &= \frac{\frac{I_{Q_s}(\theta)}{P(\theta)}\phi\left(\frac{\tilde{\theta}_1^* - \alpha_1\theta_1}{\sqrt{\sigma_{11}^2+\sigma_{21}^2}}\right)\frac{1}{\sigma_{21}^2}\phi\left(\frac{\tilde{\theta}_{21}-\tilde{\theta}_1^*/\alpha_1}{\alpha_2}\right)\prod_{k=2}^{K}\frac{1}{\sigma_{1k}}\phi\left(\frac{\hat{\theta}_{1k}-\theta_k}{\sigma_{1k}}\right)}{\frac{I_{Q_s}(\theta)}{P(\theta)}\phi\left(\frac{\alpha_1\theta_1-\theta_1^*}{\sqrt{\sigma_{11}^2+\sigma_{21}^2}}\right)\prod_{k=2}^{K}\frac{1}{\sigma_{1k}}\phi\left(\frac{\hat{\theta}_{1k}-\theta_k}{\sigma_{1k}}\right)\frac{\alpha_2}{\sigma_{21}^2}\Phi(\omega_a)} \\ &= \frac{\frac{1}{\alpha_2}\phi\left(\frac{\tilde{\theta}_{21}-\tilde{\theta}_1^*/\alpha_1}{\alpha_2}\right)}{\Phi(\omega_a)}. \end{aligned} \tag{6.18}$$

Since $\tilde{\theta}_1^* < \frac{\sigma_{21}}{\sigma_{11}}\hat{\theta}_{12} + \frac{\sigma_{11}}{\sigma_{21}}\tilde{\theta}_{21}$, the support of $\tilde{\theta}_{21}$ needed to Rao-Blackwellise $\tilde{\theta}_{21}$ is given by $A = \frac{\sigma_{21}}{\sigma_{11}}(\tilde{\theta}_1^* - \frac{\sigma_{21}}{\sigma_{11}}\hat{\theta}_{12})$. Hence, using the result of the mean of a truncated normal described in Section 3.3.2, the UMVCUE for θ_1 accounting for selection of the minimum log HR and correlation between stagewise statistics is

$$\begin{aligned} \tilde{\theta}_{1a} = E[\tilde{\theta}_{21}|\tilde{\boldsymbol{\theta}}, Q_s] &= \int_A^\infty \tilde{\theta}_{21}\frac{\frac{1}{\alpha_2}\phi\left(\frac{\tilde{\theta}_{21}-\tilde{\theta}_1^*/\alpha_1}{\alpha_2}\right)}{\Phi(\omega_a)}\mathrm{d}\tilde{\theta}_{21} \\ &= \frac{1}{\Phi(\omega_a)}\left[\frac{\tilde{\theta}_1^*}{\alpha_1} - \int_{-\infty}^{A}\frac{\tilde{\theta}_{21}}{\alpha_2}\phi\left(\frac{\tilde{\theta}_{21}-\tilde{\theta}_1^*/\alpha_1}{\alpha_2}\right)\mathrm{d}\tilde{\theta}_{21}\right] \\ &= \frac{1}{\Phi(\omega_a)}\left[\frac{\tilde{\theta}_1^*}{\alpha_1} + \alpha_2\phi(\omega_a) - \frac{\tilde{\theta}_1^*}{\alpha_1}\left(1-\Phi(\omega_a)\right)\right] \\ &= \frac{\tilde{\theta}_1^*}{\alpha_1} + \alpha_2\frac{\phi(\omega_a)}{\Phi(\omega_a)} \\ &= \frac{\sigma_{21}^2\hat{\theta}_{11} + \sigma_{11}^2\tilde{\theta}_{21}}{\sigma_{11}^2+\sigma_{21}^2} + \frac{\sigma_{21}^2}{\sqrt{\sigma_{11}^2+\sigma_{21}^2}}\frac{\phi(\omega_a)}{\Phi(\omega_a)}. \end{aligned} \tag{6.19}$$

The UMVCUE is asymptotically unbiased for the selected treatment. As separate stage 1 estimates were assumed, this estimator is of a similar form to the Bowden and Glimm estimator and the Kimani et al. estimator with no stopping for futility, but with two main differences. Firstly, this estimator considers a selection rule where the minimum log HR among all experimental treatments is selected at the interim analysis, that is, the treatment with maximum efficacy. As can be seen in

the derivation, this difference in selection rule for estimation of a log HR, assuming asymptotic normality, changes the limits of integration and thus the sign of the bias correction from negative to positive.

Secondly, this estimator now adjusts for the correlation between stagewise estimates due to censoring by defining independent increments for stage 2 data. This estimator is therefore appropriate for time-to-event data and is an extension to the Bowden and Glimm [2008] estimator where the differences are: directly estimating the treatment difference although with separate control arms, selection of the minimum log HR and accounting for correlation due to censoring.

6.8 Bias correction with a common control arm

Pairwise comparisons of each experimental treatment to a common control arm lead to correlated stage 1 estimates at the interim analysis. The UMVCUE derived in the previous section assumed separate control arms in order to maintain independence. This section therefore replaces this assumption to account for correlated stage 1 statistics, as described by density (6.12).

A recent paper by Robertson et al. [2016] gives a general framework for calculating UMVCUE's in the multivariate normal setting which explicitly accounts for the correlation between stage 1 estimates. Their setting of interest is a two-stage genome-wide association study where at the end of stage 1, estimates may be correlated. Using their general framework, the UMVCUE derived in the preceding section is now adapted to account for the correlation within stage 1.

Recall the density for $(\hat{\theta}_{11}, \hat{\theta}_{12}, ..., \hat{\theta}_{1K}, \tilde{\theta}_{21})'$ (6.12) derived in Section 6.6. Let $\mathbf{\Sigma}$ denote the variance-covariance matrix for stage 1 statistics from density (6.12). Since the stage 2 increment statistic, $\tilde{\theta}_{21}$, is unbiased for θ_1, the UMVCUE for θ_1 is found by Rao-Blackwellisation of $\tilde{\theta}_{21}$, given sufficient and complete statistics. These

are derived as follows.

Let $\hat{\boldsymbol{\theta}}$ denote the vector of stage 1 estimates $(\hat{\theta}_{11}, ..., \hat{\theta}_{1K})'$ and $\boldsymbol{\theta}$ denote the vector of true log HRs $(\theta_1, ..., \theta_K)'$. Then, following expression (6.12) the joint density of $\hat{\boldsymbol{\theta}}$ and the stage 2 independent increment, $\tilde{\theta}_{21}$, given by expression (6.12), conditional on selection, can be written as

$$f(\hat{\boldsymbol{\theta}}, \tilde{\theta}_{21}|Q_s) = \frac{I_{Q_s}(\hat{\boldsymbol{\theta}})}{P(\hat{\boldsymbol{\theta}})} g(\hat{\boldsymbol{\theta}}) \frac{1}{\sigma_{21}} \phi\left(\frac{\tilde{\theta}_{21} - \theta_1}{\sigma_{21}}\right), \tag{6.20}$$

where $I_{Q_s}(\hat{\boldsymbol{\theta}})$ is the indicator function for Q_s, $P(\hat{\boldsymbol{\theta}}) = Prob(I_{Q_s}(\hat{\boldsymbol{\theta}}) = 1)$ and

$$g(\hat{\boldsymbol{\theta}}) \quad = \quad \frac{1}{\sqrt{(2\pi)^K |\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right\}.$$

Now in order to determine the sufficient and complete statistic for $\boldsymbol{\theta}$, the exponent of the joint density (6.20) can be written as

$$\frac{1}{2}\left[2\left(\sum_{k=1}^{K} \sigma_{1,1,k}\hat{\theta}_{1k} + \frac{1}{\sigma_{21}^2}\tilde{\theta}_{21}\right)\theta_1 + 2\sum_{k=2}^{K}\left(\sum_{l=1}^{K}\sigma_{1,k,l}\hat{\theta}_{1l}\right)\theta_k - \boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta} - \frac{1}{\sigma_{21}^2}\theta_1^2 + \psi(\hat{\boldsymbol{\theta}}, \tilde{\theta}_{21})\right],$$

where $\psi(\hat{\boldsymbol{\theta}}, \tilde{\theta}_{21}) = -\hat{\boldsymbol{\theta}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\theta}} - \frac{1}{\sigma_{21}^2}\tilde{\theta}_{21}^2$. Thus from the above density and as shown in Appendix B, by Theorem 3.4.6 and Definition 3.4.8,

$$\tilde{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}} + \frac{\boldsymbol{\Sigma}_1}{\sigma_{21}^2}\tilde{\theta}_{21}, \tag{6.21}$$

is sufficient and complete for $\boldsymbol{\theta}$ where $\boldsymbol{\Sigma}_1 = (\sigma_{1,1,1}, ..., \sigma_{1,1,K})'$. Now by Rao-Blackwellisation of the stage 2 increment, the UMVCUE is derived by $\mathrm{E}[\tilde{\theta}_{21}|\tilde{\boldsymbol{\theta}}^*, Q_s]$. Thus, we seek the conditional density $f(\tilde{\theta}_{21}|\tilde{\boldsymbol{\theta}}^*, Q_s)$, which is found by deriving $f(\tilde{\theta}_{21}, \tilde{\boldsymbol{\theta}}^*|Q_s)$ and $f(\tilde{\boldsymbol{\theta}}^*|Q_s)$.

From the definition of $\tilde{\boldsymbol{\theta}}^*$, we can write

$$\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}^* - \frac{\boldsymbol{\Sigma}_1}{\sigma_{21}^2}\tilde{\theta}_{21}. \tag{6.22}$$

Therefore, density $f(\hat{\boldsymbol{\theta}}, \tilde{\theta}_{21}|Q_s)$ (6.20) can be re-expressed as

$$\begin{aligned} f(\tilde{\theta}_{21}, \tilde{\boldsymbol{\theta}}^*|Q_s) &= \frac{I_{Q_s}(\hat{\boldsymbol{\theta}})}{P(\hat{\boldsymbol{\theta}})} g\left(\tilde{\boldsymbol{\theta}}^* - \frac{\boldsymbol{\Sigma}_1}{\sigma_{21}^2}\tilde{\theta}_{21}\right) \frac{1}{\sigma_{21}}\phi\left(\frac{\tilde{\theta}_{21} - \theta_1}{\sigma_{21}}\right) \\ &= \frac{I_{Q_s}(\hat{\boldsymbol{\theta}})}{P(\hat{\boldsymbol{\theta}})} \frac{1}{\sqrt{2\pi\alpha_2^2}} \exp\left\{-\frac{1}{2\alpha_2^2}\left(\tilde{\theta}_{21} - \frac{\sigma_{21}}{\sigma_{11}}\theta_1^*/\alpha_1\right)^2\right\}, \end{aligned} \tag{6.23}$$

where α_1 and α_2 are as defined in Section 6.7.

The density $f(\tilde{\boldsymbol{\theta}}^*|Q_s)$, also needed to derive the UMVCUE, is found by integrating density $f(\tilde{\theta}_{21}, \tilde{\boldsymbol{\theta}}^*|Q_s)$ (6.23) over the support of $\tilde{\theta}_{21}$. As we condition on selection of the minimum log HR, the support of $\tilde{\theta}_{21}$ is found as follows.

The definition of Q_s implies $\hat{\theta}_{1k} < \hat{\theta}_{1k+1}$ for $k = 1, ..., K-1$. Thus from equation (6.22)

$$\begin{aligned} \hat{\theta}_{1k} < \hat{\theta}_{1k+1} &\implies \tilde{\theta}_k^* - \frac{\sigma_{1,1,k}}{\sigma_{21}^2}\tilde{\theta}_{21} < \tilde{\theta}_{k+1}^* - \frac{\sigma_{1,1,k+1}}{\sigma_{21}^2}\tilde{\theta}_{21} \\ &\implies \sigma_{21}^2(\tilde{\theta}_k^* - \tilde{\theta}_{k+1}^*) < \tilde{\theta}_{21}(\sigma_{1,1,k} - \sigma_{1,1,k+1}). \end{aligned}$$

Hence, if $\sigma_{1,1,k} > \sigma_{1,1,k+1}$ then $\tilde{\theta}_{21} > \frac{\sigma_{21}^2(\tilde{\theta}_k^* - \tilde{\theta}_{k+1}^*)}{\sigma_{1,1,k} - \sigma_{1,1,k+1}}$. Conversely, if $\sigma_{1,1,k} < \sigma_{1,1,k+1}$ then $\tilde{\theta}_{21} < \frac{\sigma_{21}^2(\tilde{\theta}_k^* - \tilde{\theta}_{k+1}^*)}{\sigma_{1,1,k} - \sigma_{1,1,k+1}}$. However, if $\sigma_{1,1,k} = \sigma_{1,1,k+1}$ then there is no restriction on the integral with respect to $\tilde{\theta}_{21}$. Putting all of these inequalities together gives $c_1 < \tilde{\theta}_{21} < c_2$, where

$$c_1 = \begin{cases} \max\left\{\frac{\sigma_{21}^2(\tilde{\theta}_1^* - \tilde{\theta}_2^*)}{\sigma_{11}^2 - \sigma_{1,1,2}}, ..., \frac{\sigma_{21}^2(\tilde{\theta}_{K-1}^* - \tilde{\theta}_K^*)}{\sigma_{1,1,K-1} - \sigma_{1,1,K}}\right\}, & \text{if } \sigma_{1,1,k} > \sigma_{1,1,k+1}, \\ -\infty, & \text{otherwise,} \end{cases} \tag{6.24}$$

and

$$c_2 = \begin{cases} \min\left\{ \dfrac{\sigma_{21}^2(\tilde{\theta}_1^* - \tilde{\theta}_2^*)}{\sigma_{11}^2 - \sigma_{1,1,2}}, ..., \dfrac{\sigma_{21}^2(\tilde{\theta}_{K-1}^* - \tilde{\theta}_K^*)}{\sigma_{1,1,K-1} - \sigma_{1,1,K}} \right\}, & \text{if } \sigma_{1,1,k} < \sigma_{1,1,k+1}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{6.25}$$

Hence, from equations (6.23), (6.24) and (6.25) the expression for the UMVCUE, $\tilde{\theta}_{1b}$, is

$$\begin{aligned} \mathrm{E}[\tilde{\theta}_{21}|\tilde{\boldsymbol{\theta}}^*, Q_s] &= \frac{\displaystyle\int_{c_1}^{c_2} \tilde{\theta}_{21} f(\tilde{\theta}_{21}, \tilde{\boldsymbol{\theta}}^*|Q_s) \mathrm{d}\tilde{\theta}_{21}}{\displaystyle\int_{c_1}^{c_2} f(\tilde{\boldsymbol{\theta}}^*|Q_s) \mathrm{d}\tilde{\theta}_{21}} \\ &= \frac{\displaystyle\int_{c_1}^{c_2} \tilde{\theta}_{21} \frac{1}{\sqrt{2\pi\alpha_2^2}} \exp\left\{ -\frac{1}{2\alpha_2^2} \left(\tilde{\theta}_{21} - \frac{\sigma_{21}}{\sigma_{11}} \tilde{\theta}_1^* / \alpha_1 \right)^2 \right\} \mathrm{d}\tilde{\theta}_{21}}{\displaystyle\int_{c_1}^{c_2} \frac{1}{\sqrt{2\pi\alpha_2^2}} \exp\left\{ -\frac{1}{2\alpha_2^2} \left(\tilde{\theta}_{21} - \frac{\sigma_{21}}{\sigma_{11}} \tilde{\theta}_1^* / \alpha_1 \right)^2 \right\} \mathrm{d}\tilde{\theta}_{21}}. \end{aligned} \tag{6.26}$$

Now using the result in equation (3.9) and following steps in Section 3.3.5, expression (6.26) is evaluated to give the asymptotic UMVCUE for time-to-event data. This estimator corrects for selection of the minimum log HR and both within- and between-stage correlation and is given by

$$\tilde{\theta}_{1b} = \frac{\sigma_{21}^2 \hat{\theta}_{11} + \sigma_{11}^2 \tilde{\theta}_{21}}{\sigma_{11}^2 + \sigma_{21}^2} + \frac{\sigma_{21}^2}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}} \frac{\phi(\omega_{c_1}) - \phi(\omega_{c_2})}{\Phi(\omega_{c_2}) - \Phi(\omega_{c_1})}, \tag{6.27}$$

where $\omega_{c_i} = \dfrac{c_i \sqrt{\sigma_{11}^2 + \sigma_{21}^2}}{\sigma_{21}^2} - \dfrac{\tilde{\theta}_1^*}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}}$ for $i = 1, 2$.

The form of this estimator is similar to the Robertson et al. [2016] estimator, which corrects for selection of the maximum treatment mean and correlation between stage 1 estimates. However, this new estimator corrects for treatment selection in the opposite direction, that is, selection of the minimum log HR. Additionally, this estimator corrects for both the correlation between stage 1 estimates and the correlation between stage 1 and 2 estimates for the selected treatment.

The estimators derived in this chapter are summarised in Table 6.2.

Table 6.2: Summary of estimators for the selected log HR, θ_1

Estimator	Equation	Description
$\hat{\theta}_1$	-	Naive estimator from all accrued data at the final analysis that ignores selection and both correlation within and between stages.
$\tilde{\theta}_{21}$	(6.10)	Stage 2 increment estimate for the selected treatment.
$\tilde{\theta}_{1a}$	(6.19)	Asymptotic UMVCUE that accounts for selection and correlation between stages but ignores correlation within stage 1.
$\tilde{\theta}_{1b}$	(6.27)	Asymptotic UMVCUE that accounts for selection and both correlation between stages and within stage 1.

6.8.1 Bivariate asymptotic UMVCUE

For the case where $K = 2$, the UMVCUE given in expression (6.27) can be simplified to

$$\tilde{\theta}_{1b} = \begin{cases} \dfrac{\sigma_{21}^2\hat{\theta}_{11} + \sigma_{11}^2\tilde{\theta}_{21}}{\sigma_{11}^2 + \sigma_{21}^2} + \dfrac{\sigma_{21}^2}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}} \dfrac{\phi(\omega_b)}{\Phi(-\omega_b)}, & \text{if } \sigma_{11}^2 > \sigma_{1,1,2}, \\ \dfrac{\sigma_{21}^2\hat{\theta}_{11} + \sigma_{11}^2\tilde{\theta}_{21}}{\sigma_{11}^2 + \sigma_{21}^2} - \dfrac{\sigma_{21}^2}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}} \dfrac{\phi(\omega_b)}{\Phi(\omega_b)}, & \text{if } \sigma_{11}^2 < \sigma_{1,1,2}, \\ \dfrac{\sigma_{21}^2\hat{\theta}_{11} + \sigma_{11}^2\tilde{\theta}_{21}}{\sigma_{11}^2 + \sigma_{21}^2}, & \text{if } \sigma_{11}^2 = \sigma_{1,1,2}, \end{cases} \quad (6.28)$$

where $\omega_b = \dfrac{(\tilde{\theta}_1^* - \tilde{\theta}_2^*)\sqrt{\sigma_{11}^2 + \sigma_{21}^2}}{\sigma_{11}^2 - \sigma_{1,1,2}} - \dfrac{\tilde{\theta}_1^*}{\sqrt{\sigma_{11}^2 + \sigma_{21}^2}}$, $\tilde{\theta}_1^* = \hat{\theta}_{11} + \dfrac{\sigma_{11}^2}{\sigma_{21}^2}\tilde{\theta}_{21}$ and $\tilde{\theta}_2^* = \hat{\theta}_{12} + \dfrac{\sigma_{1,1,2}}{\sigma_{21}^2}\tilde{\theta}_{21}$.

6.9 Simulation study

6.9.1 Simulation set up

The properties of the stage 2 increment estimator, $\tilde{\theta}_{21}$, the naive estimator, $\hat{\theta}_1$, and the UMVCUE, $\tilde{\theta}_{1b}$, are now compared in a simulation study. Since we are concerned with unbiased and efficient estimation, their bias and MSE are assessed for the influence of the number of experimental treatment arms, treatment effect sizes and selection times. As in the motivational simulation study (Section 6.3), up to 6 experimental treatments are considered so that $K \in \{1, ..., 6\}$ with selection time defined by $\tau \in \{0.1, ..., 0.9\}$. Various scenarios for different effect sizes are considered: scenario (a) assumes all experimental treatments are equally ineffective, where log HRs $\theta_k = 0$ $(k = 1, ..., K)$; (b) assumes all treatments are equally effective, where $\theta_k = -0.22$ $(k = 1, ..., K)$; and (c) assumes one uniquely superior treatment, where $\theta_1 = -0.22$ and $\theta_k = 0$ $(k = 2, ..., K)$. Additionally, unequal effect sizes are assumed for all treatment arms to assess a range of effect sizes from small to large log HRs. Here, log HRs $\theta_k = -0.05k$ $(k = 1, ..., K)$, which correspond to HRs from 0.78 to 0.95.

This simulation study is set up to resemble a seamless phase II/III clinical trial with patients randomised by blocked randomisation with block size $2(K + 1)$. This means that two patients are allocated to each experimental treatment and control arm within each block. Randomisation times are simulated from 0 to 6 months from a uniform distribution for a total of $n = 1500$ patients per group. Event times E_j are simulated for each patient $j = 1, ..., n$ from a Weibull distribution using the R package by Stasinopoulos et al. [2006], with shape parameter $\gamma = 1.5$ for all arms and a range of rate parameters $\lambda_c \in \{0.15, 0.25\}$ and $\lambda_k \in \{0.03, 0.1, 0.12, 0.15\}$ for the control arm c and experimental treatment arms k. Random censoring times C_j are simulated for each patient $j = 1, ..., n$ for each arm from an exponential distribution with

rate 0.001. A patients survival time is thus calculated as $\min\{E_j, C_j\}$. The survival times assumed here correspond to true log HRs $\theta_k \in \{-1.61, -0.92, -0.51, -0.22, 0\}$. The main assumption here is that the distribution of event times is stochastically independent of the censoring time distribution.

As discussed, for time-to-event data, the information fraction depends on the number of events. Therefore, the total number of events at the final analysis is set to $D = 900$, such that the total number of events at the interim analysis is $d = \tau D$. Additionally, the length of follow-up for patients recruited in stage 1 with censored events at the interim analysis is limited to T_{follow}, as described in Section 6.4. Thus, the additional number of events from these patients is defined as $D_{\text{follow}} = d + 0.1D$.

In each simulated trial, at the end of stage 1, an interim analysis is conducted where log HRs, $\hat{\theta}_{1k}$, are estimated based on the log-rank score statistic and variance as described in Section 6.6 and equations (6.1) and (6.2). The observed effect sizes for each treatment are ranked in order of decreasing magnitude and the treatment associated with the smallest log HR is selected to take forward into stage 2. At the end of stage 2, after a total of 900 events, including delayed events followed-up until T_{follow}, a final analysis is conducted where the estimators of interest given in Table 6.2 are computed.

A naive estimator, $\hat{\theta}_1$, is computed as described in Section 6.3. This estimator does not account for selection or correlation, and is obtained at the final analysis based on all accrued data for the selected treatment and control arm. The stage 2 increment, $\tilde{\theta}_{21}$, is computed as described by equation (6.10). This increment estimate is the difference between the estimate from all trial data, where delayed events from stage 1, which occur after T_{follow} are censored at T_{follow}, and the estimate computed at the interim analysis for the selected treatment. The asymptotic UMVCUE is computed using expressions (6.27), (6.24), (6.25) and (6.21). In order to compute the sufficient

statistics $\tilde{\boldsymbol{\theta}}^*$, the covariance of any two ordered stage 1 statistics is obtained as described by equation (6.11).

The average selection bias and RMSE are calculated from 10,000 simulated trials for each value of K and 9 values of τ in the interval (0.1,0.9). For each trial, selection bias of each estimator for the selected treatment is calculated as the mean of the differences in the estimated log HR and the true log HR. The MSE is calculated as the mean of the squared differences between the estimator and the true log HR. Recall, since $\theta_k < 0$ $(k = 1, ..., K)$ implies superiority of treatment k, a negative bias implies a positively biased estimator, that is, the estimator overestimates the true treatment effect.

6.9.2 Simulation results

Figures 6.4 shows the bias (top plots) and RMSE (bottom plots) of the stage 2 increment as a function of selection time for scenarios (a)-(c). The bias and RMSE for asymptotic UMVCUE are given in Figure 6.5. For the case of equally ineffective treatments, that is, when $\theta_k = 0$ $(k = 1, ..., K)$, it can be seen that the stage 2 increment and the new UMVCUE are asymptotically unbiased for all selection times and for any number of experimental treatment arms, up to six. The stage 2 increment has a larger RMSE compared to the UMVCUE. Specifically, when selection is made later in the trial, it can be seen that the RMSE of the stage 2 increment increases steeply with selection time. This reduction in efficiency is expected since this estimator is the difference between the final estimate and the interim estimate. The UMVCUE, on the other hand, has a much improved RMSE for all values of K, with a small increase observed for late selection times of $\tau > 0.8$. Hence, the UMVCUE has better properties than the stage 2 increment estimator.

In comparison to the simulation study of the naive estimator (Figure 6.2), the new

UMVCUE corrects for the overestimation of treatment effects in the naive estimator for all selection times and any number of experimental treatment arms. Specifically, Figure 6.6 shows a comparison of all estimators for the case of 2, 4 and 6 experimental treatment arms. In terms of RMSE, the UMVCUE has a larger RMSE compared to the naive estimator. This was also shown for the normal case by Kimani et al. [2013], where they observed a larger RMSE for their UMCUE compared to the naive estimator for all selection times. As can be seen the difference in RMSE's decreases with the number of experimental treatment arms (Figure 6.6 (bottom)).

In Section 6.3, it was shown that the naive estimator is positively biased in all cases with bias increasing with the number of experimental treatment arms. In particular, for the case of equally effective treatments (scenario (b)), overestimation of up to -0.04 was observed for the naive estimator. The new UMVCUE shows a considerable improvement in this bias for all selection times compared to the naive estimator (Figure 6.5(b)). For selection times of $\tau < 0.8$, bias appears to improve with the number of experimental treatment arms K. However, for late selection times $\tau > 0.8$, the UMVCUE slightly underestimates the true effect for trials with a small number of treatment comparisons ($K \leq 3$) and slightly overestimates for trials with a large number of treatment comparisons ($K \geq 6$) (Figure 6.7). This implies that the UMVCUE provides a more conservative estimate of the selected treatment effect in cases with a small number of treatment comparisons for late selection times. However, this is still a considerable improvement in bias over the naive estimator and, as discussed in Section 6.3, this is due to the inherent bias of the asymptotic normality assumption.

In terms of RMSE, as was observed for the case of equally ineffective treatments, the UMVCUE has a larger RMSE, when all treatments are equally effective, compared to the naive estimator for all values of τ and K, with a gradual increase observed with selection time up to $\tau = 0.8$.

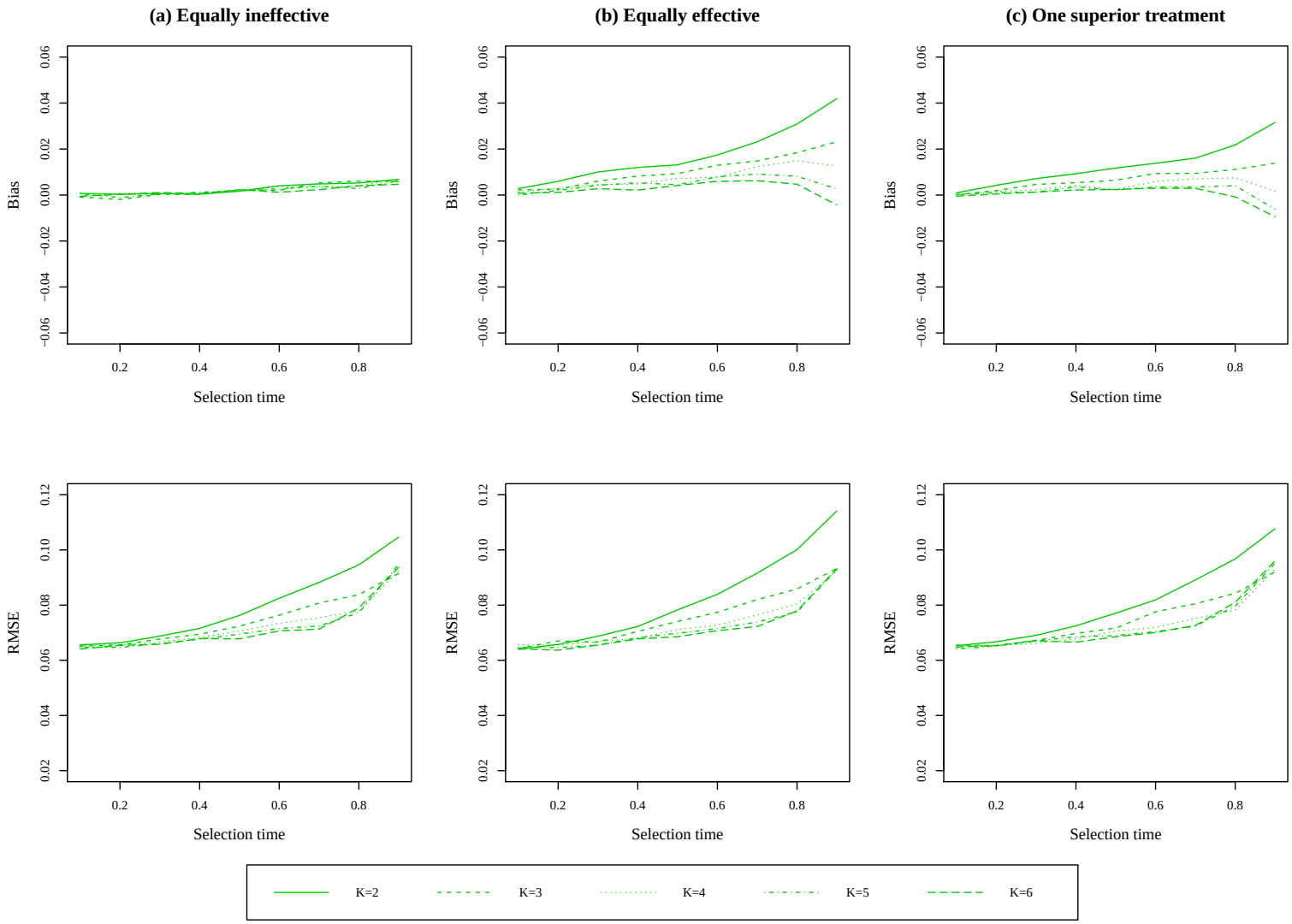

Figure 6.4: Bias (top row) and RMSE (bottom row) of the stage 2 increment estimator $\tilde{\theta}_{21}$ for $K \in \{2, ..., 6\}$ treatment groups and selection times $\tau \in \{0.1, ..., 0.9\}$ from 10,000 simulations. Each column represents a different combination of effect sizes: (a) $\theta_k = 0 \ \forall k \in \{1, ..., K\}$, (b) $\theta_k = -0.22 \ \forall k \in \{1, ..., K\}$, and (c) $\theta_1 = -0.22$ and $\theta_k = 0 \ \forall k \in \{2, ..., K\}$.

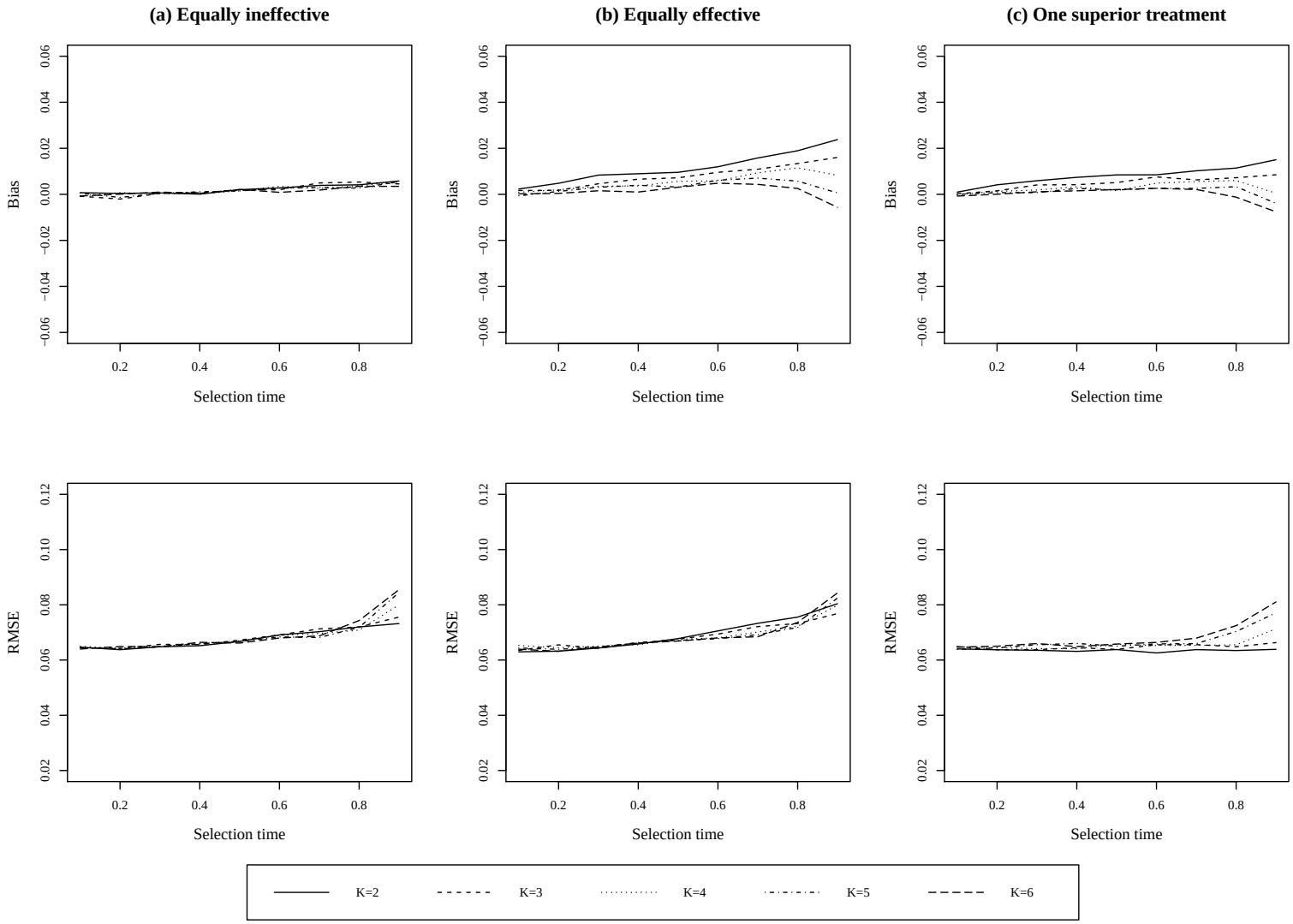

Figure 6.5: Bias (top row) and RMSE (bottom row) of the asymptotic UMVCUE for $K \in \{2, ..., 6\}$ treatment groups and selection times $\tau \in \{0.1, ..., 0.9\}$ from 10,000 simulations. Each column represents a different combination of effect sizes: (a) $\theta_k = 0 \; \forall k \in \{1, ..., K\}$, (b) $\theta_k = -0.22 \; \forall k \in \{1, ..., K\}$, and (c) $\theta_1 = -0.22$ and $\theta_k = 0 \; \forall k \in \{2, ..., K\}$.

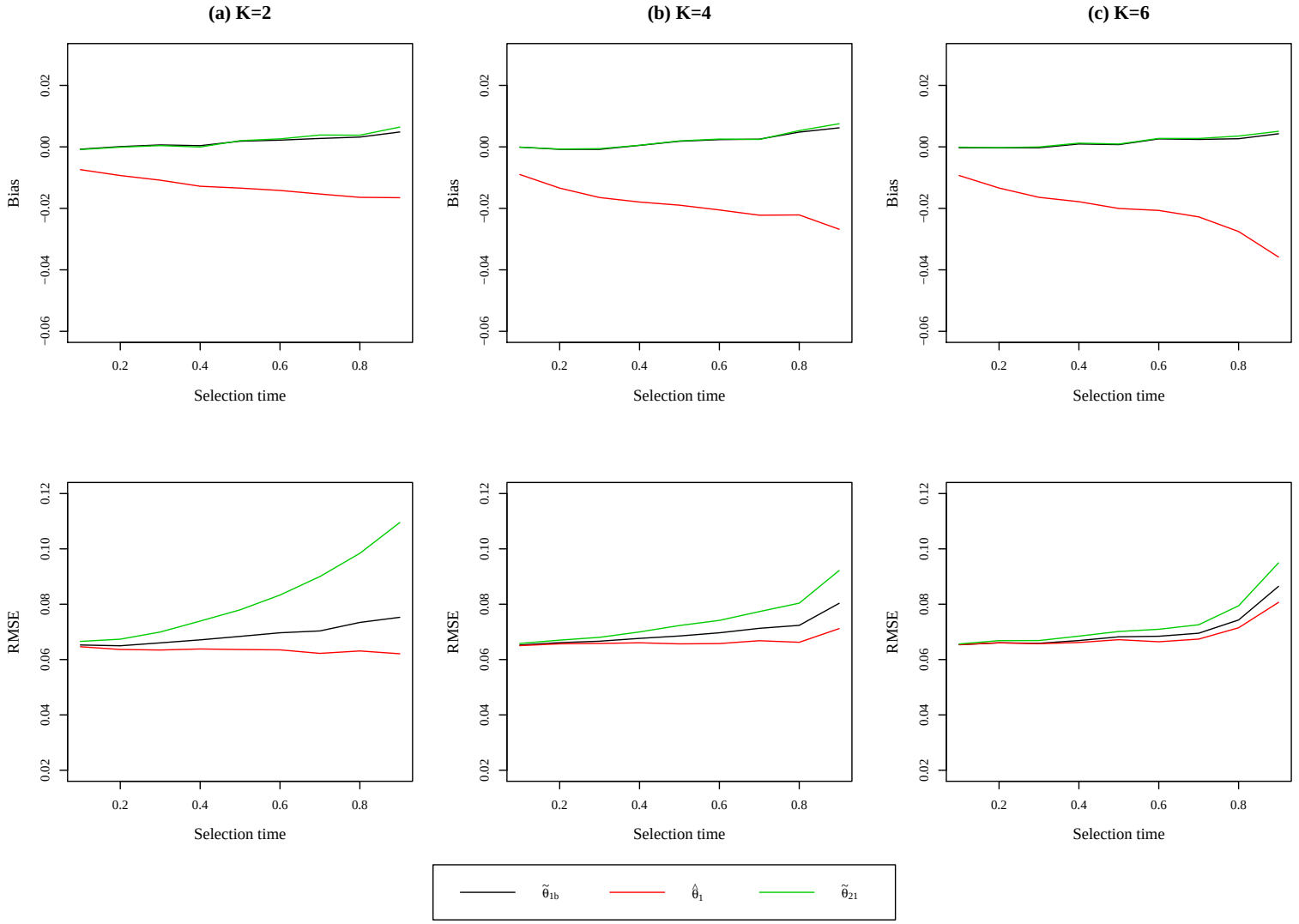

Figure 6.6: Bias (top row) and RMSE (bottom row) of the naive estimator, stage 2 increment estimator, and asymptotic UMVCUE for a different number of treatment groups K and selection times $\tau \in \{0.1, ..., 0.9\}$, from 10,000 simulations. For each experimental treatment arm k, log HRs $\theta_k = 0 \; \forall k \in \{1, ..., K\}$.

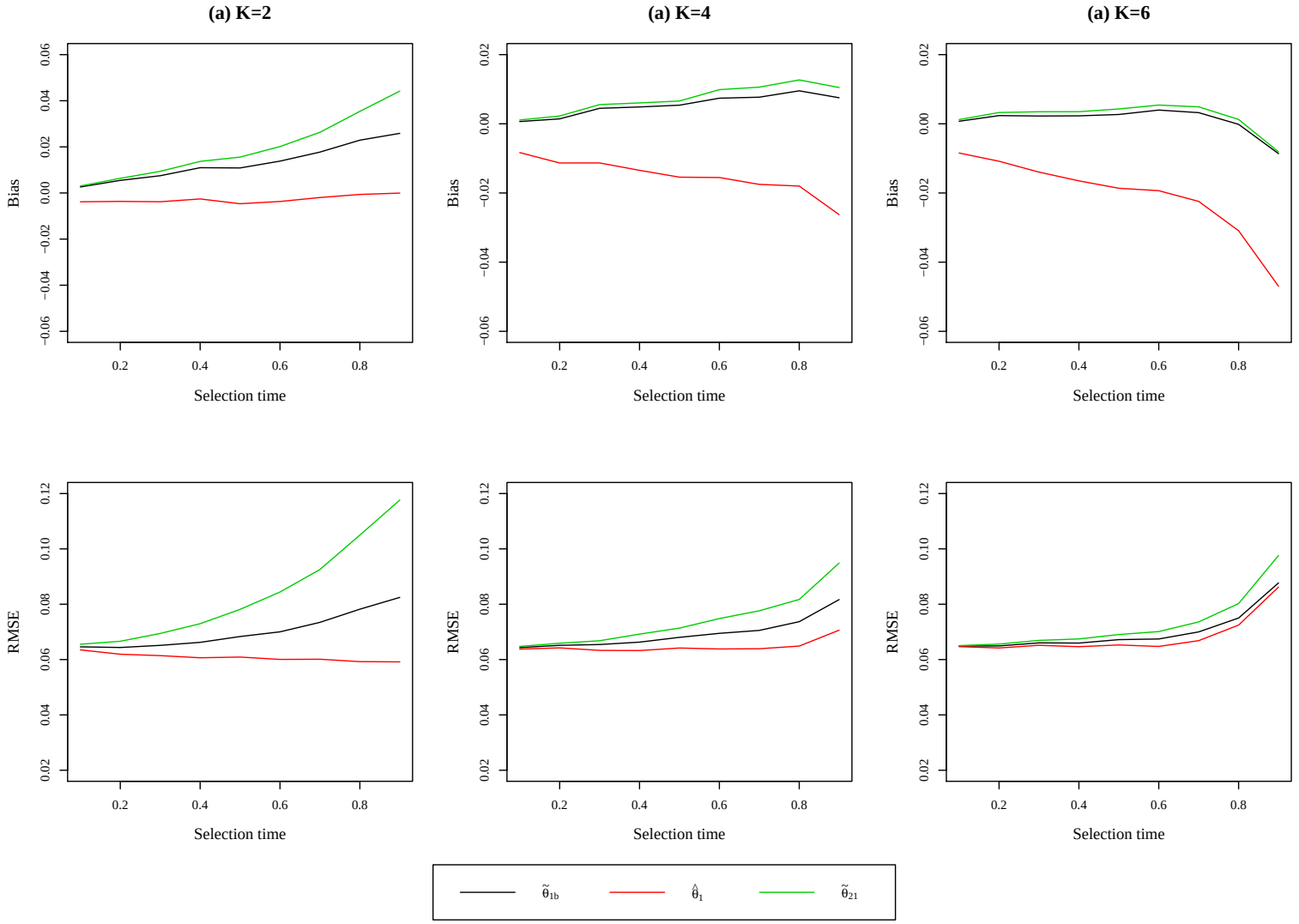

Figure 6.7: Bias (top row) and RMSE (bottom row) of the naive estimator, stage 2 increment estimator, and asymptotic UMVCUE for a different number of treatment groups K and selection times $\tau \in \{0.1, ..., 0.9\}$, from 10,000 simulations. For each experimental treatment arm k, log HRs $\theta_k = -0.22 \; \forall k \in \{1, ..., K\}$.

A similar result is observed for the case of one superior treatment, as can be seen in Figure 6.5(c). However, the magnitude of bias is overall smaller than the case of equally effective treatments with $\theta_k = 0.8$ $(k = 1, ..., K)$. This is expected since we have seen previously that bias reduces as the difference in the selected treatment effect and second best treatment effect increases.

For all scenarios, the UMVCUE RMSE appears to increase steeply for late selection times $\tau > 0.8$. Additionally, for the case of one superior treatment, the RMSE increases with the number of experimental treatments for selection times $\tau > 0.8$. This is similarly observed for the stage 2 increment, with a steeper increase across selection times. This is partly due to the inherent bias from the normality assumption, but also due to the correlation between stage 1 and 2 data, as later selection times correspond to a larger information fraction which means a greater proportion of events from patients recruited in stage 1 are included in the final analysis.

Despite the new UMVCUE having a slightly larger RMSE compared with the naive estimator, these simulations have shown a substantial improvement in bias over the naive estimator for all selection times. These results therefore show that the UMVCUE has considerably better properties than the naive estimator, since it corrects for the substantial bias that arises from both treatment selection and correlation due to censored observations. However, as the UMVCUE is only asymptotically unbiased, where the normality approximation of the log-rank score statistics are only appropriate for small to moderate treatment effects, it is important to consider the treatment effect size in the bias and RMSE trade-off.

Therefore, we now assess the influence of the effect size on bias and RMSE. We first assume equal log HRs for all experimental treatment arms in the range -1.61 to 0. Log HRs $\theta_k \leq -0.9$ are considered large effect sizes, while $-0.9 < \theta_k < 0$ are considered moderate to small effect sizes. Table 6.3 shows the bias and RMSE of the naive estimator, $\hat{\theta}_1$, the stage 2 increment estimate, $\tilde{\theta}_{21}$, and the UMVCUE, $\tilde{\theta}_{1b}$,

for selection times $\tau = (0.3, 0.5, 0.8)$ for the case of $K = 4$ experimental treatment arms. For large effect sizes, it can be seen that all estimators underestimate the true treatment effect, with the UMVCUE performing the worst. Additionally, Figure 6.8(a) indicates the UMVCUE bias increases with effect size and selection time, where a bias of up to 0.3 is observed for a large log HR ($\theta_k = -1.61$). However, in a comparison of Figures 6.8(a)-(c), bias appears to reduce as the number of experimental treatments increase. Therefore, this indicates that for large HRs the UMVCUE does not correct for the bias due to treatment selection, as we have previously observed that selection bias in the naive estimator is influenced by the number of experimental treatment arms. Furthermore, a considerable difference in RMSE for $\theta_k = -1.61$ and $\theta_k > -1.61$ can be seen in all cases of K. These results, therefore, highlight that the asymptotic approximation of the log HR does not hold for large treatment effects. Hence, due to both the asymptotic assumptions as well as bias due to selection of large log HRs, the UMVCUE is only appropriate for small to moderate effect sizes.

Now to assess the effect of varying log HRs, we assume true effect sizes $\theta_k = -0.05k$ ($k = 1, ..., K$), which correspond to HRs from 0.78 to 0.95. Figure 6.9 shows the bias (top row) and RMSE (bottom row) of the naive estimator, $\hat{\theta}_1$, the stage 2 increment estimate, $\tilde{\theta}_{21}$, and the UMVCUE, $\tilde{\theta}_{1b}$, for selection times $\tau \in \{0.1, ..., 0.9\}$ and $K = \{1, ..., 6\}$ experimental treatment arms. The UMVCUE corrects for the overestimation of treatment effects in the naive estimator, as was similarly observed for the normal case of unequal means in Chapter 5. As can be seen, the UMVCUE is asymptotically unbiased for all selection times for $K \leq 4$. However, for late selection times $\tau > 0.8$, all estimators indicate a small increase in bias for increasing number of experimental treatments. In terms of RMSE, the stage 2 increment has a larger RMSE compared to the naive estimator and UMVCUE, as expected. In particular, the RMSE increases more steeply as both selection time increases and the number of experimental treatments reduces. In contrast, the UMVCUE has an improved

Table 6.3: Mean bias and RMSE from 10,000 simulations of the naive estimator ($\hat{\theta}_1$), the stage 2 increment ($\tilde{\theta}_{21}$), and the UMVCUE ($\tilde{\theta}_{1b}$) for the case of two experimental treatment arms at selection times $\tau = \{0.3, 0.5, 0.8\}$. Log HRs θ_k $(k = 1, ..., 4)$ are assumed to be equal for each scenario.

			True effect size				
			θ_k:	0	-0.22	-0.51	-0.92	-1.61
	τ	Estimator	HR:	1	0.8	0.6	0.4	0.2
Bias	0.3	$\hat{\theta}_1$		-0.0145	-0.0113	-0.0047	0.0180	0.1345
		$\tilde{\theta}_{21}$		0.0004	0.0056	0.0138	0.0374	0.1460
		$\tilde{\theta}_{1b}$		0.0006	0.0045	0.0118	0.0350	0.1493
	0.5	$\hat{\theta}_1$		-0.0168	-0.0154	-0.0087	0.0147	0.1372
		$\tilde{\theta}_{21}$		0.0034	0.0066	0.0168	0.0422	0.1610
		$\tilde{\theta}_{1b}$		0.0032	0.0054	0.0132	0.0375	0.1603
	0.8	$\hat{\theta}_1$		-0.0224	-0.0180	-0.0088	0.0104	0.1220
		$\tilde{\theta}_{21}$		0.0043	0.0127	0.0279	0.0528	0.1739
		$\tilde{\theta}_{1b}$		0.0041	0.0095	0.0210	0.0428	0.1608
RMSE	0.3	$\hat{\theta}_1$		0.0641	0.0633	0.0625	0.0635	0.1463
		$\tilde{\theta}_{21}$		0.0665	0.0668	0.0693	0.0771	0.1602
		$\tilde{\theta}_{1b}$		0.0649	0.0654	0.0670	0.0743	0.1618
	0.5	$\hat{\theta}_1$		0.0654	0.0642	0.0623	0.0625	0.1481
		$\tilde{\theta}_{21}$		0.0716	0.0714	0.0744	0.0843	0.1760
		$\tilde{\theta}_{1b}$		0.0683	0.0680	0.0696	0.0783	0.1731
	0.8	$\hat{\theta}_1$		0.0669	0.0649	0.0616	0.0616	0.1345
		$\tilde{\theta}_{21}$		0.0802	0.0817	0.0869	0.1007	0.1957
		$\tilde{\theta}_{1b}$		0.0725	0.0737	0.0770	0.0875	0.1788

RMSE, with only a gradual increase observed across selection times. However, the RMSE is overall larger for all values of K compared to the naive estimator, as was observed in previous scenarios and for the normal case.

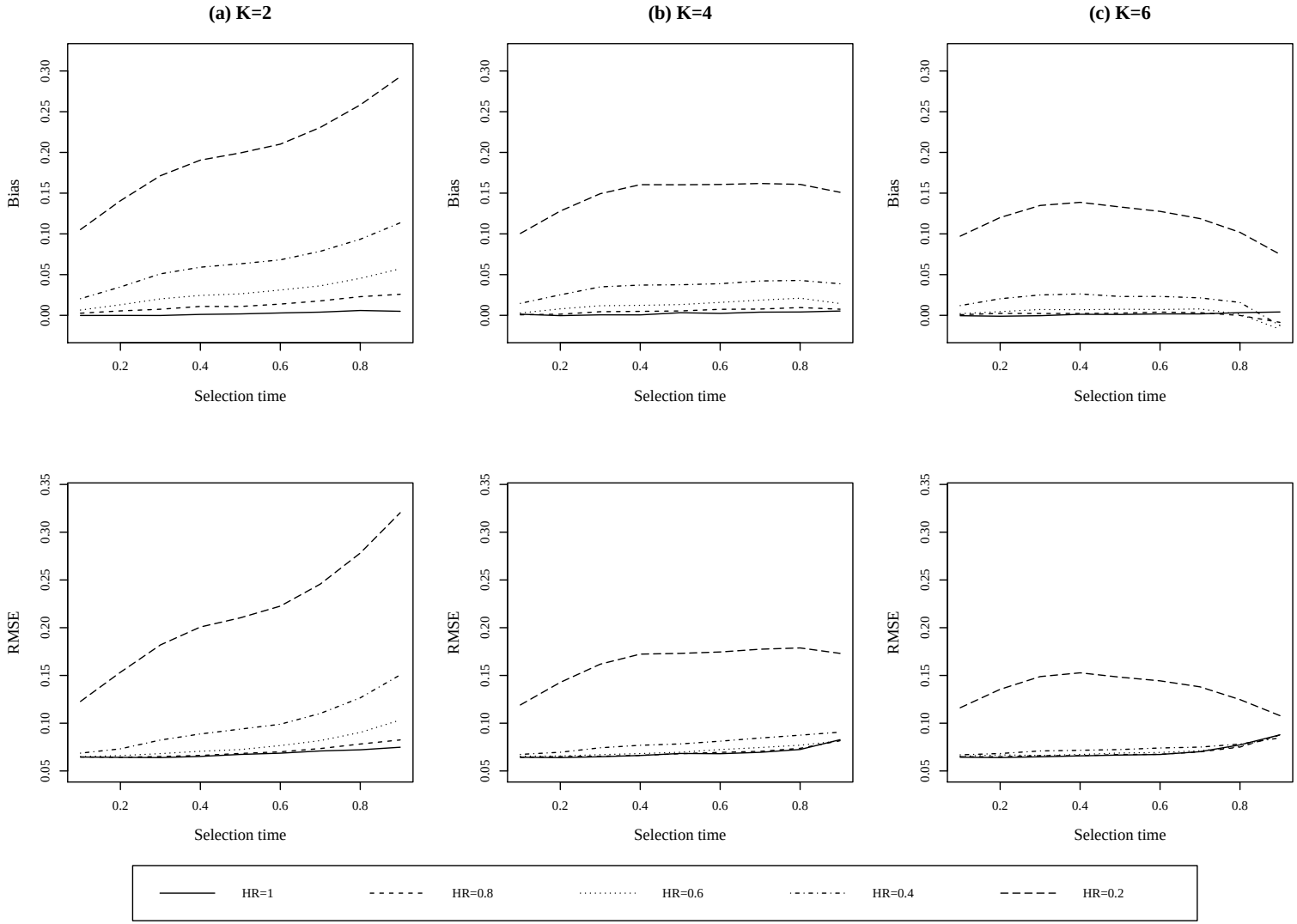

Figure 6.8: Bias (top row) and RMSE (bottom row) of the asymptotic UMVCUE for a range of effect sizes from 10,000 simulations. A range of HRs (0.2,1) are assumed equal for K treatment arms and selection times $\tau \in \{0.1, ..., 0.9\}$.

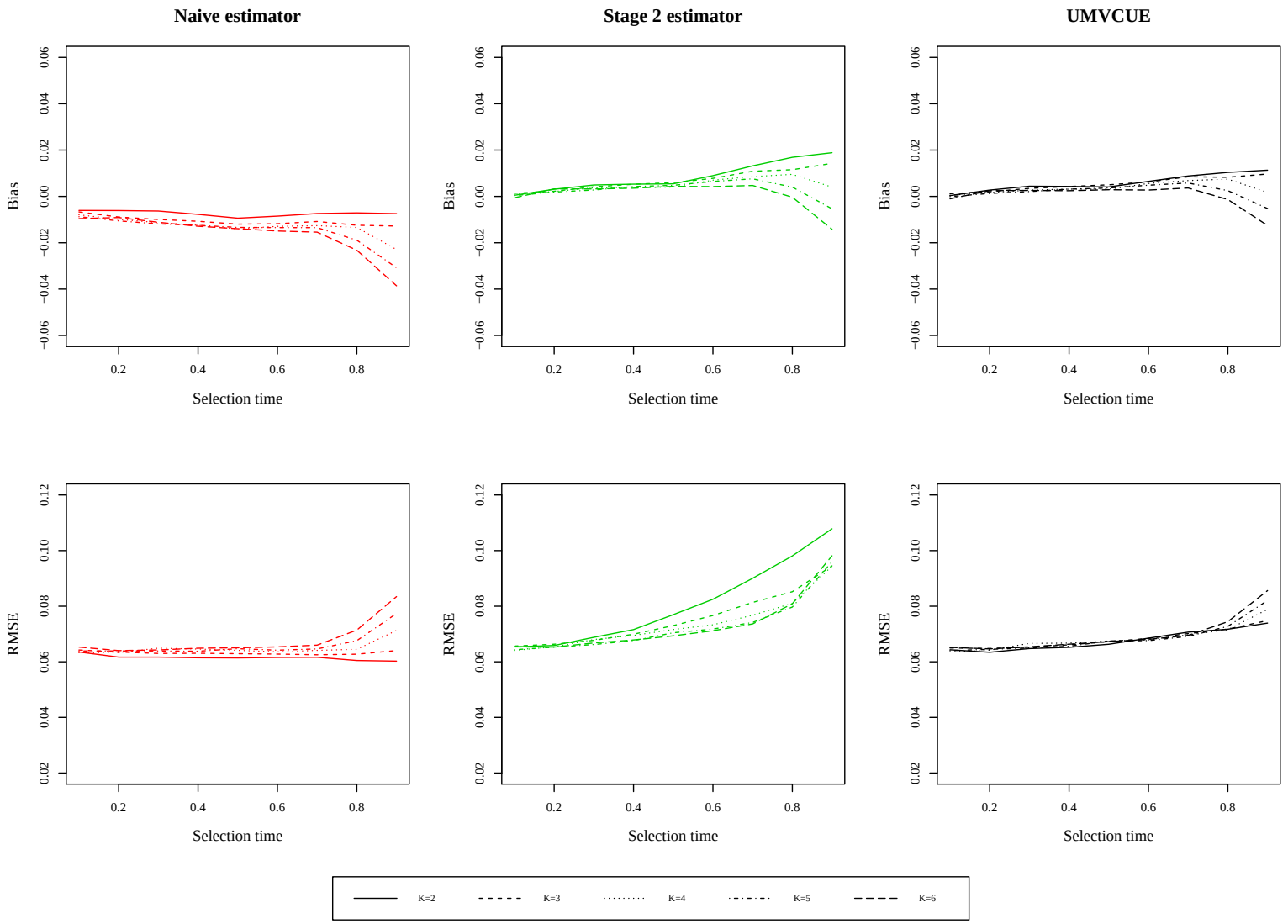

Figure 6.9: Bias (top row) and RMSE (bottom row) of the naive estimator, stage 2 increment estimator, and asymptotic UMVCUE for unequal effect sizes for each treatment arm from 10,000 simulations. For each experimental treatment arm k, log HRs $\theta_k = -0.05k \; \forall k \in \{1, ..., K\}$.

6.10 Conclusion

This chapter addressed the importance and need for unbiased and efficient estimators for time-to-event data in seamless phase II/III clinical trials. The problem of selection bias arises when data are used for both selection of the most effective treatment at the interim stage, as well as estimation of the treatment effect at the final analysis. Methods of estimation that account for this problem use the technique of Rao-Blackwellisation of the unbiased but inefficient stage 2 data. However, these methods only exist for normally distributed data and are thus not appropriate for censored survival data. Therefore, by extension of these methods, this chapter developed two asymptotically unbiased estimators that correct for selection bias as well as correlation of stagewise estimates due to censored observations at the interim analysis.

Independent increment statistics were introduced in order to decorrelate stagewise statistics and thus derive the joint asymptotic distribution of stage 1 and stage 2 log HRs. The first estimator assumed separate control arms in order to decorrelate stage 1 estimates. An asymptotic unbiased estimator correcting for selection bias and correlation between stages was derived by extension of the ideas from Bowden and Glimm [2008], Di Scala and Glimm [2011] and Kimani et al. [2013]. However, the main limitation of this estimator is that it assumes a separate control group for each experimental treatment group. In practice, this may not be a realistic assumption as investigators and regulators are generally interested in saving resources, such as costs and the number of patients recruited in a trial.

Therefore, since a common control arm is generally considered in practice, the second estimator derived in this chapter replaces this assumption to include a common control arm for all experimental treatment arms. This however leads to correlated stage 1 estimates due to the common control arm. Thus, extending the previous es-

timator with the methods given by Robertson et al. [2016] for a multivariate normal setting, the second UMVCUE developed corrects for both the correlation between stage 1 estimates as well as the correlation between stages, while also adjusting for treatment selection bias for time-to-event data.

Properties of the UMVCUE with a common control arm were compared to the naive estimator and stage 2 increment estimator in a simulation study for various trial scenarios. Under the null, the UMVCUE was found to be asymptotically unbiased for any selection time and any number of experimental treatments. Furthermore, a substantial improvement in bias was observed in the UMVCUE compared to the naive estimator for small to moderate effect sizes, with bias improving with the number of experimental treatment arms. However, for large effect sizes, such as a HR of 0.2, results highlighted that the asymptotic assumptions of the log HR fail to hold. Therefore, the UMVCUE has favourable properties over the naive estimator relative to the treatment effect size and the desired bias-RMSE trade-off.

Chapter 7

Subpopulation selection in adaptive survival trials

7.1 Setting

In targeted therapy trials, a biomarker is used to identify a subset of patients with certain molecular or gene expressions. It is thought that this subgroup or subpopulation may respond better to treatment than the full population [Buyse et al., 2011]. Figure 7.1 illustrates the population in a given trial, where the full population, F, is made up of the subpopulation of interest, defined as the biomarker-positive group, S, and its complement S^c, defined as the biomarker-negative group. This setting is common in targeted therapy trials and is therefore the setting assumed in this chapter.

Statistical analysis of such data encounters multiplicity issues as treatment efficacy is assessed in the biomarker-positive subgroup as well as the full population in order to identify the population that benefits from the new intervention. As discussed throughout this thesis, two-stage adaptive trials which combine the learning and

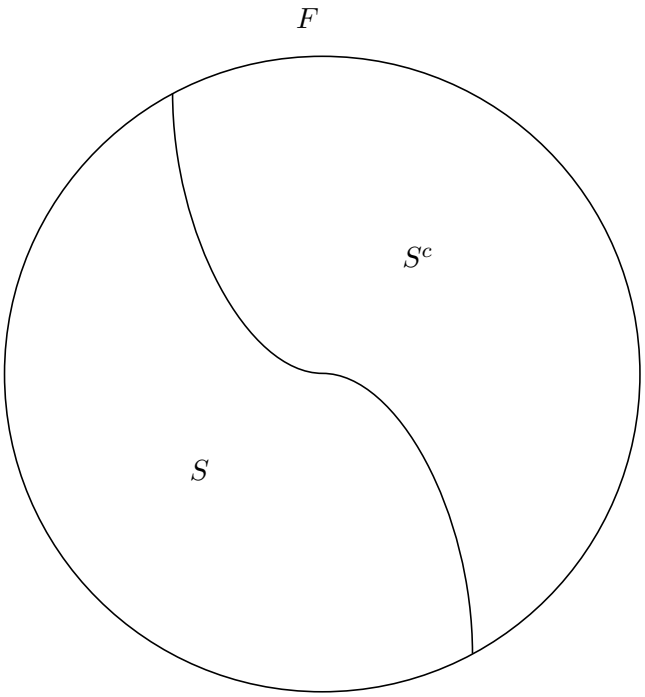

Figure 7.1: Population diagram. The full population is denoted by F, and the biomarker-positive and biomarker-negative subpopulations are denoted by S and S^c, respectively.

confirming phases of drug development may be more efficient in addressing multiple clinical questions within one trial. In the setting of subpopulation selection, two-stage designs allow both selection of the subpopulation to be made in an independent learning phase, as well as testing of treatment efficacy in a confirmatory phase III setting. This is an attractive design where both phases are combined into a single two-stage trial as illustrated in Figure 7.2.

In stage 1, randomisation to the control arm or experimental treatment arm is stratified by the prespecified biomarker subpopulations S and S^c to ensure an equal number of patients are assigned to each arm within each subpopulation. At the interim analysis, the decision of whether to continue the trial to stage 2 with either the biomarker-positive subpopulation, S, or the full population, F, is based on a predefined selection criterion. If the treatment is deemed sufficiently effective in the biomarker-positive subpopulation, then trial continues with S only. On the other

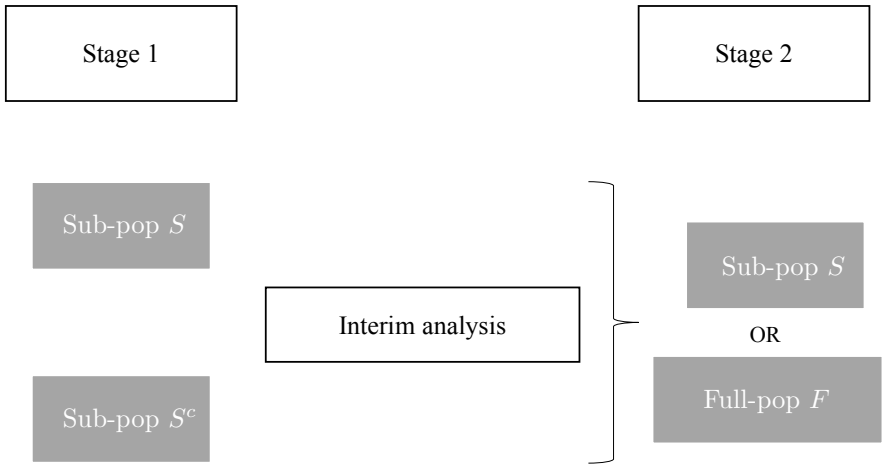

Figure 7.2: Two stage trial schema with subpopulation selection

hand, if the treatment effect in the subpopulation is not sufficiently better than the effect in the full population, then the trial continues with the full population. This is described in more detail in Section 7.3.

Although this design is beneficial in combining two phases of drug development into a single trial, statistical issues arise with treatment effect estimation and multiple hypothesis testing. A chance selection of the most promising population may occur at the interim analysis which may lead to bias in the naive estimate from stage 1 data. Hence, at the end of the trial, the final effect estimate may be overestimated and thus biased as stage 1 data are used for both population selection as well as estimation of treatment efficacy. Several papers, including Friede et al. [2012], Hommel [2001] and Glimm and Di Scala [2015], propose methods that address multiplicity which arises in hypothesis testing for normally distributed data in this setting. For time-to-event data, papers including Brannath et al. [2009], Jenkins et al. [2011] and Mehta et al. [2014] address the issues of multiplicity for various trial settings with subpopulation selection at the interim. Specifically, Jenkins et al. [2011] propose using prespecified weights in order to utilise all trial data and thus ensure inde-

pendence of stagewise statistics for time-to-event data. Additionally, Glimm and Di Scala [2015] suggest their method holds under the assumptions of asymptotic normality of the log-rank test statistic or the Cox regression coefficient of the treatment effect. All of these methods, however, focus on statistical testing of the null hypothesis rather than estimation of the treatment effect.

In terms of point estimation, for normally distributed outcomes, Kimani et al. [2015] propose conditionally unbiased estimators correcting for subpopulation selection bias in two-stage trials. They derive UMVCUEs for treatment means of the selected population, S or F. For time-to-event outcomes, point estimators that account for subpopulation selection have not been developed. Therefore, in this chapter we combine the methods from Kimani et al. with those developed in the previous chapter in order to derive asymptotically conditionally unbiased estimators. We assume the setting of subpopulation selection in Kimani et al. [2015], but for time-to-event data.

7.1.1 Sources of bias

In two-stage two-arm survival trials, as mentioned before, bias may be introduced in the naive estimator as data from all patients recruited up until the interim analysis time are used for both subpopulation selection as well as estimation of the treatment effect at the final analysis.

In addition, censoring at the interim analysis leads to correlated stagewise test statistics due to the continued follow-up in stage 2 of patients recruited in stage 1. As described in the previous chapter, such data are referred to as delayed events. Thus, bias may depend on the additional follow-up of stage 1 patients with delayed events. In order to ensure the information in stage 2 does not depend on data observed from patients recruited in stage 1, the length of additional follow-up time

for such delayed events should be prefixed and unrelated to the selection rule, as discussed in Section 6.4.

7.2 Notation

Consider a two-stage two-arm trial with one experimental treatment compared to a control in two subpopulations denoted by $k = \{S, S^c\}$. Let p_k denote the known sample prevalence of subpopulation k from the full population F. Suppose n_i patients are recruited in each stage i $(i = 1, 2)$ and let $n_{ik} = p_k \cdot n_i$ denote the number of patients in subpopulation k in stage i, where recruitment is constrained in each subpopulation. Then the total number of patients in stage 1 is $n_1 = n_{1S} + n_{1S^c}$ and stage 2 is

$$n_2 = \begin{cases} n_{2S} & \text{if } S \text{ is selected} \\ n_{2S} + n_{2S^c} & \text{if } F \text{ is selected.} \end{cases}$$

Key notation for this chapter is summarised in Table 7.1.

7.2.1 Estimating the log HRs

As interest is in the survival distributions within two subpopulations S and S^c, separate log HRs are estimated for the treatment effect within each subgroup. Let θ_k denote the true log HR for the experimental treatment in subpopulation k $(k = S, S^c)$ and $\hat{\theta}_{1k}$ denote the estimated log HR for subpopulation k in stage 1. If S is selected, let $\hat{\theta}_{2S}$ denote the estimated log HR at the final analysis, and if F is selected, let $\hat{\theta}^F_{2k}$ denote the estimated log HR for subpopulation k at the final analysis.

As before, stagewise log HRs are computed based on the log-rank score statistic and variance by $\hat{\theta}_{1k} = \frac{S_{1k}}{V_{1k}}$, $\hat{\theta}_{2S} = \frac{S_{2S}}{V_{2S}}$ and $\hat{\theta}^F_{2k} = \frac{S^F_{2k}}{V^F_{2k}}$, where S_{ik}, V_{ik}, S^F_{2k} and V^F_{2k}

Table 7.1: Key notation

Selected population	Subpopulation	Stage 1	Stage 2	Stages 1 and 2		
				Naive estimator	Sufficient statistic	Unbiased estimator
S	S	$\hat{\theta}_{1S} \sim N(\theta_S, \sigma^2_{1S})$	$\tilde{\theta}_{2S} \sim N(\theta_S, \sigma^2_{2S})$	$\hat{\theta}_S$	$\tilde{\theta}^*_S$	$\tilde{\theta}_S$
	S^c	$\hat{\theta}_{1S^c} \sim N(\theta_{S^c}, \sigma^2_{1S^c})$	-	-	-	-
F	S	$\hat{\theta}_{1S} \sim N(\theta_S, \sigma^2_{1S})$	$\tilde{\theta}^F_{2S} \sim N(\theta_S, \sigma^2_{2S^F})$	$\hat{\theta}^F_S$	$\tilde{\theta}^*_{S^F}$	$\tilde{\theta}^F_S$
	S^c	$\hat{\theta}_{1S^c} \sim N(\theta_{S^c}, \sigma^2_{1S^c})$	$\tilde{\theta}^F_{2S^c} \sim N(\theta_{S^c}, \sigma^2_{2S^cF})$	$\hat{\theta}^F_{S^c}$	$\tilde{\theta}^*_{S^c}$	$\tilde{\theta}^F_{S^c}$

$(i = 1, 2)$ are as defined by equations (6.1) and (6.2). Hence, $\hat{\theta}_{1k}$ is asymptotically normal with mean θ_k and variance $\sigma^2_{1k} = \frac{1}{V_{1k}}$.

As discussed in Section 6.6, to ensure independence between stagewise test statistics for time-to-event outcomes, independent increments are utilised with a prespecified additional follow-up time for stage 1 delayed events. If S is selected, as the stage 2 estimator, $\hat{\theta}_{2S}$, is based on all trial data, the stage 2 independent increment is defined by

$$\tilde{\theta}_{2S} = \frac{S_{2S} - S_{1S}}{V_{2S} - V_{1S}}, \tag{7.1}$$

which is asymptotically normal with mean θ_S and variance $\sigma^2_{2S} = \frac{1}{V_{2S} - V_{1S}}$.

On the other hand, if F is selected, then independent increments are defined for each subpopulation k $(k = S, S^c)$ by

$$\tilde{\theta}^F_{2k} = \frac{S^F_{2k} - S_{1k}}{V^F_{2k} - V_{1k}}, \tag{7.2}$$

which is asymptotically normal with mean θ_k and variance $\sigma^2_{2k^F} = \frac{1}{V^F_{2k} - V_{1k}}$.

7.3 Subpopulation selection rule

A treatment is deemed most efficacious if it is one with the smallest observed log HR. In terms of subpopulation selection, at the interim analysis the trial continues with subpopulation S if the experimental treatment is observed to be sufficiently more effective in S than in F. Subpopulation S is selected if $\hat{\theta}_{1S} < \hat{\theta}_{1S^c} - c$, where c is chosen depending on the subpopulation prevalence. Let Q denote the condition

of selection such that

$$Q = \begin{cases} \hat{\theta}_{1S} < \hat{\theta}_{1S^c} - c & \text{if } S \text{ is selected} \\ \hat{\theta}_{1S} \geq \hat{\theta}_{1S^c} - c & \text{if } F \text{ is selected.} \end{cases} \tag{7.3}$$

The selection margin c may be thought of as a penalty for small subpopulations, where the smaller the subpopulation, the less likely it is to be selected. For example, c may be defined by $c = \frac{b}{p_S}$, since dividing by the prevalence ensures that S is only selected if a strong effect exists. The numerator b may be derived, for example, using an economic gain function defined by the number of QUALY gains for selecting the subpopulation S [Ondra et al., 2016].

7.4 Case where S is selected

First consider the case where the experimental treatment is sufficiently more effective in subpopulation S compared to the full population F, such that the trial continues to the second stage with S.

7.4.1 Naive estimator for θ_S

As defined in Section 7.2.1, at the interim and final analysis, estimators $\hat{\theta}_{1S}$ and $\hat{\theta}_{2S}$ are obtained based on the log-rank score statistics from all data accrued up until each analysis time. Conditional on selection, $\hat{\theta}_{1S} \sim N(\theta_S, \frac{1}{V_{1S}})$ and $\hat{\theta}_{2S} \sim N(\theta_S, \frac{1}{V_{2S}})$. Using independent increments to decorrelate stagewise estimates, we can derive the joint distribution of stage 1 statistics and the stage 2 increment statistic for the selected subpopulation. Recall stage 1 estimates are uncorrelated since separate subgroups are considered. Therefore, following Section 6.6 the joint asymptotic

distribution is given by

$$\begin{pmatrix} \hat{\theta}_{1S} \\ \hat{\theta}_{1S^c} \\ \tilde{\theta}_{2S} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_S \\ \theta_{S^c} \\ \theta_S \end{pmatrix}, \begin{pmatrix} \frac{1}{V_{1S}} & 0 & 0 \\ 0 & \frac{1}{V_{1S^c}} & 0 \\ 0 & 0 & \frac{1}{V_{2S} - V_{1S}} \end{pmatrix} \right). \tag{7.4}$$

A naive estimator for the treatment effect at the final analysis can be defined as the weighted sum of the stage 1 log HR and the stage 2 increment statistic, weighted by the information in each stage.

Let τ_S denote the information fraction of stage 1 data for subpopulation S, such that $\tau_S = \frac{V_{1S}}{V_{2S}}$. Thus, in the case where S is selected, the naive estimator is defined by

$$\hat{\theta}_S = \tau_S \hat{\theta}_{1S} + (1 - \tau_S)\tilde{\theta}_{2S}. \tag{7.5}$$

At the final analysis, the naive estimator overestimates the true treatment effect when selection is made with respect to subpopulations. Hence, the resulting estimator, $\hat{\theta}_S$, is positively biased.

7.4.2 An asymptotically UMVCUE for θ_S

This section addresses the issue of bias due to subpopulation selection by deriving a UMVCUE, which corrects for selection bias based on the asymptotic normality assumption of the log-rank score statistic. This estimator is based on Kimani et al. [2015]. Since the independent increments property is utilised, this estimator also accounts for the correlation which arises due to censored events at the interim analysis. Furthermore, as was discussed in the previous chapter, this assumption of an independent covariance structure only holds as long as test statistics are unchanged.

Therefore in order to ensure independence is satisfied, the maximum follow-up time for censored stage 1 events is limited and prespecified, as described in Section 6.4. Additionally, due to the asymptotic assumptions, we note this estimator is only asymptotically unbiased for large samples and small effect sizes.

Recall from equation (7.3) if the subpopulation S is selected then $Q = \hat{\theta}_{1S} < \hat{\theta}_{1S^c} - c$. Thus, conditional on selection and following density (7.4), the joint density of the stage 1 log HRs and the stage 2 increment for the selected subpopulation is

$$f(\hat{\theta}_{1S}, \hat{\theta}_{1S^c}, \tilde{\theta}_{2S}|Q) = \frac{I_{[Q(\theta)]}}{P(\theta)} \frac{1}{\sigma_{2S}} \phi\left(\frac{\tilde{\theta}_{2S} - \theta_S}{\sigma_{2S}}\right) \frac{1}{\sigma_{1S}} \phi\left(\frac{\hat{\theta}_{1S} - \theta_S}{\sigma_{1S}}\right) \frac{1}{\sigma_{1S^c}} \phi\left(\frac{\hat{\theta}_{1S^c} - \theta_{S^c}}{\sigma_{1S^c}}\right) \tag{7.6}$$

where $I_{[Q(\theta)]}$ is the indicator function for Q and $P(\theta) = Prob(I_{[Q(\theta)]} = 1)$.

Since the increment $\tilde{\theta}_{2S}$ is unbiased for θ_S, the UMVCUE is derived using the method of Rao-Blackwellisation (Theorem 3.5.3). Therefore, the required complete, sufficient statistics for θ_S are now derived as follows.

Following the result in expression (3.11), the above joint density is re-expressed as

$$f(\hat{\theta}_{1S}, \hat{\theta}_{1S^c}, \tilde{\theta}_{2S}|Q) = \frac{I_{[Q(\theta)]}}{P(\theta)} \phi\left(\frac{\left(\frac{\sigma_{2S}}{\sigma_{1S}}\hat{\theta}_{1S} + \frac{\sigma_{1S}}{\sigma_{2S}}\tilde{\theta}_{2S}\right) - \alpha_1\theta_S}{\sqrt{\sigma_{1S}^2 + \sigma_{2S}^2}}\right) \frac{1}{\sigma_{1S^c}} \phi\left(\frac{\hat{\theta}_{1S^c} - \theta_{S^c}}{\sigma_{1S^c}}\right) \frac{1}{\sigma_{1S}\sigma_{2S}} \phi\left(\frac{\hat{\theta}_{1S} - \left(\frac{\sigma_{2S}}{\sigma_{1S}}\hat{\theta}_{1S} + \frac{\sigma_{1S}}{\sigma_{2S}}\tilde{\theta}_{2S}\right)/\alpha_1}{\frac{\sigma_{1S}^2}{\sigma_{2S}^2}\alpha_2}\right) \tag{7.7}$$

where $\alpha_1 = \frac{\sigma_{1S}}{\sigma_{2S}} + \frac{\sigma_{2S}}{\sigma_{1S}}$ and $\alpha_2 = \frac{\sigma_{2S}^2}{\sqrt{\sigma_{1S}^2+\sigma_{2S}^2}}$. Let $\tilde{\theta}_S^* = \frac{\sigma_{2S}}{\sigma_{1S}}\hat{\theta}_{1S} + \frac{\sigma_{1S}}{\sigma_{2S}}\tilde{\theta}_{2S}$, then by transformation of random variables, density $f(\hat{\theta}_{1S}, \hat{\theta}_{1S^c}, \tilde{\theta}_{2S}|Q)$ can be transformed

to $f(\hat{\theta}_{1S}, \hat{\theta}_{1S^c}, \tilde{\theta}^*_S|Q)$ with the Jacobian being $\frac{\sigma_{2S}}{\sigma_{1S}}$, to give the density

$$f(\hat{\theta}_{1S}, \hat{\theta}_{1S^c}, \tilde{\theta}^*_S|Q) = \frac{I_{[Q(\theta)]}}{P(\theta)}\phi\left(\frac{\tilde{\theta}^*_S - \alpha_1\theta_S}{\sqrt{\sigma^2_{1S}+\sigma^2_{2S}}}\right)\frac{1}{\sigma^2_{1S}}\phi\left(\frac{\hat{\theta}_{1S} - \tilde{\theta}^*_S/\alpha_1}{\frac{\sigma^2_{1S}}{\sigma^2_{2S}}\alpha_2}\right) \frac{1}{\sigma_{1S^c}}\phi\left(\frac{\hat{\theta}_{1S^c} - \theta_{S^c}}{\sigma_{1S^c}}\right). \tag{7.8}$$

Hence, by the Factorisation Theorem 3.4.6, from the above density it follows that $(\tilde{\theta}^*_S, \hat{\theta}_{1S^c})$ is minimal sufficient and thus a complete and sufficient statistic for θ_S. Now, to derive the UMVCUE we need to find the expression for $E[\tilde{\theta}_{2S}|\hat{\theta}_{1S^c}, \tilde{\theta}^*_S, Q]$. Since $E[\tilde{\theta}_{2S}|\hat{\theta}_{1S^c}, \tilde{\theta}^*_S, Q] = \int \tilde{\theta}_{2S} f(\tilde{\theta}_{2S}|\hat{\theta}_{1S^c}, \tilde{\theta}^*_S, Q)$, by standard theory of conditional probabilities we seek the densities $f(\tilde{\theta}_{2S}, \hat{\theta}_{1S^c}, \tilde{\theta}^*_S|Q)$ and $f(\hat{\theta}_{1S^c}, \tilde{\theta}^*_S|Q)$. The latter density is found by solving the integral $\int f(\hat{\theta}_{1S}, \hat{\theta}_{1S^c}, \tilde{\theta}^*_S|Q)\mathrm{d}\hat{\theta}_{1S}$. From the condition of selection, the range of $\hat{\theta}_{1S}$ is $(-\infty, \hat{\theta}_{1S^c} - c)$. Hence,

$f(\hat{\theta}_{1S^c}, \tilde{\theta}^*_S|Q)$

$$\begin{aligned}
&= \int_{-\infty}^{\hat{\theta}_{1S^c}-c} f(\hat{\theta}_{1S}, \hat{\theta}_{1S^c}, \tilde{\theta}^*_S|Q)\mathrm{d}\hat{\theta}_{1S} \\
&= \frac{I_{[Q(\theta)]}}{P(\theta)}\frac{1}{\sigma^2_{1S}}\phi\left(\frac{\tilde{\theta}^*_S - \alpha_1\theta_S}{\sqrt{\sigma^2_{1S}+\sigma^2_{2S}}}\right)\frac{1}{\sigma_{1S^c}}\phi\left(\frac{\hat{\theta}_{1S^c} - \theta_{S^c}}{\sigma_{1S^c}}\right)\int_{-\infty}^{\hat{\theta}_{1S^c}-c}\phi\left(\frac{\hat{\theta}_{1S} - \tilde{\theta}^*_S/\alpha_1}{\frac{\sigma^2_{1S}}{\sigma^2_{2S}}\alpha_2}\right)\mathrm{d}\hat{\theta}_{1S} \\
&= \frac{I_{[Q(\theta)]}}{P(\theta)}\phi\left(\frac{\tilde{\theta}^*_S - \alpha_1\theta_S}{\sqrt{\sigma^2_{1S}+\sigma^2_{2S}}}\right)\frac{1}{\sigma_{1S^c}}\phi\left(\frac{\hat{\theta}_{1S^c} - \theta_{S^c}}{\sigma_{1S^c}}\right)\frac{\alpha_2}{\sigma^2_{2S}}\Phi(\omega_S)
\end{aligned}$$

where

$$\omega_S = \frac{\sqrt{\sigma^2_{1S}+\sigma^2_{2S}}}{\sigma^2_{1S}}\left\{\hat{\theta}_{1S^c} - c - \frac{\sigma^2_{2S}\hat{\theta}_{1S} + \sigma^2_{1S}\tilde{\theta}_{2S}}{\sigma^2_{1S}+\sigma^2_{2S}}\right\}.$$

Now to find the density $f(\tilde{\theta}_{2S}, \hat{\theta}_{1S^c}, \tilde{\theta}^*_S|Q)$, by following the steps described in section

3.3.5, density (7.6) is re-expressed as

$$f(\hat{\theta}_{1S}, \hat{\theta}_{1S^c}, \tilde{\theta}_{2S}|Q) = \frac{I_{[Q(\theta)]}}{P(\theta)}\phi\left(\frac{\left(\frac{\sigma_{2S}}{\sigma_{1S}}\hat{\theta}_{1S} + \frac{\sigma_{1S}}{\sigma_{2S}}\tilde{\theta}_{2S}\right) - \alpha_1\theta_S}{\sqrt{\sigma_{1S}^2 + \sigma_{2S}^2}}\right)\frac{1}{\sigma_{1S^c}}\phi\left(\frac{\hat{\theta}_{1S^c} - \theta_{S^c}}{\sigma_{1S^c}}\right)$$
$$\frac{1}{\sigma_{1S}\sigma_{2S}}\phi\left(\frac{\tilde{\theta}_{2S} - \left(\frac{\sigma_{2S}}{\sigma_{1S}}\hat{\theta}_{1S} + \frac{\sigma_{1S}}{\sigma_{2S}}\tilde{\theta}_{2S}\right)/\alpha_1}{\alpha_2}\right).$$

Thus, by transformation of random variables (Section 3.1.1) density $f(\hat{\theta}_{1S}, \hat{\theta}_{1S^c}, \tilde{\theta}_{2S}|Q)$ is transformed to $f(\tilde{\theta}_{2S}, \hat{\theta}_{1S^c}, \tilde{\theta}_S^*|Q)$, where the Jacobian of the transformation is $\frac{\sigma_{1S}}{\sigma_{2S}}$. This gives the density $f(\tilde{\theta}_{2S}, \hat{\theta}_{1S^c}, \tilde{\theta}_S^*|Q)$ as

$$\frac{I_{[Q(\theta)]}}{P(\theta)}\phi\left(\frac{\tilde{\theta}_S^* - \alpha_1\theta_S}{\sqrt{\sigma_{1S}^2 + \sigma_{2S}^2}}\right)\frac{1}{\sigma_{2S}^2}\phi\left(\frac{\tilde{\theta}_{2S} - \tilde{\theta}_S^*/\alpha_1}{\alpha_2}\right)\frac{1}{\sigma_{1S^c}}\phi\left(\frac{\hat{\theta}_{1S^c} - \theta_{S^c}}{\sigma_{1S^c}}\right).$$

Hence, the conditional density of $\tilde{\theta}_{2S}$ given complete, sufficient statistics required to derive the UMVCUE is

$$\begin{aligned} f(\tilde{\theta}_{2S}|\hat{\theta}_{1S^c}, \tilde{\theta}_S^*, Q) &= \frac{f(\tilde{\theta}_{2S}, \hat{\theta}_{1S^c}, \tilde{\theta}_S^*|Q)}{f(\hat{\theta}_{1S^c}, \tilde{\theta}_S^*|Q)} \\ &= \frac{\frac{1}{\alpha_2}\phi\left(\frac{\tilde{\theta}_{2S} - \tilde{\theta}_S^*/\alpha_1}{\alpha_2}\right)}{\Phi(\omega_S)}. \end{aligned} \tag{7.9}$$

By the condition of selection $\hat{\theta}_{1S} < \hat{\theta}_{1S^c} - c$, the support of $\tilde{\theta}_{2S}$ is equal to $L = \frac{\sigma_{2S}}{\sigma_{1S}}\left(\tilde{\theta}_S^* - \frac{\sigma_{2S}}{\sigma_{1S}}\left(\hat{\theta}_{1S^c} - c\right)\right)$. Hence, using the result of a truncated normal density

(expression (3.9)), the UMVCUE, $\tilde{\theta}_S$, is found by

$$\begin{aligned}
E[\tilde{\theta}_{2S}|\hat{\theta}_{1S^c}, \tilde{\theta}^*_S, Q] &= \int_L^\infty \tilde{\theta}_{2S} f(\tilde{\theta}_{2S}|\hat{\theta}_{1S^c}, \tilde{\theta}^*_S, Q)\mathrm{d}\tilde{\theta}_{2S} \\
&= \int_L^\infty \tilde{\theta}_{2S} \frac{\frac{1}{\alpha_2}\phi\left(\frac{\tilde{\theta}_{2S}-\tilde{\theta}^*_S/\alpha_1}{\alpha_2}\right)}{\Phi(\omega_S)} d\tilde{\theta}_{2S} \\
&= \frac{1}{\Phi(\omega_S)}\left[\frac{\tilde{\theta}^*_S}{\alpha_1} - \int_L^\infty \frac{\tilde{\theta}_{2S}}{\alpha_2}\phi\left(\frac{\tilde{\theta}_{2S}-\tilde{\theta}^*_S/\alpha_1}{\alpha_2}\right) d\tilde{\theta}_{2S}\right] \\
&= \frac{1}{\Phi(\omega_S)}\left[\frac{\tilde{\theta}^*_S}{\alpha_1} - \left[-\alpha_2\phi(-\omega_S) + \frac{\tilde{\theta}^*_S}{\alpha_1}\Phi(-\omega_S)\right]\right] \\
&= \frac{\tilde{\theta}^*_S}{\alpha_1} + \alpha_2\frac{\phi(-\omega_S)}{\Phi(\omega_S)} \\
&= \frac{\sigma^2_{2S}\hat{\theta}_{1S} + \sigma^2_{1S}\tilde{\theta}_{2S}}{\sigma^2_{1S}+\sigma^2_{2S}} + \frac{\sigma^2_{2S}}{\sqrt{\sigma^2_{1S}+\sigma^2_{2S}}}\frac{\phi(\omega_S)}{\Phi(\omega_S)}.
\end{aligned}$$

Hence, the UMVCUE when subpopulation S is selected is

$$\tilde{\theta}_S = \hat{\theta}_S + \frac{\sigma^2_{2S}}{\sqrt{\sigma^2_{1S}+\sigma^2_{2S}}}\frac{\phi(\omega_S)}{\Phi(\omega_S)}. \tag{7.10}$$

7.5 Case where F is selected

If at the interim analysis, the treatment in the biomarker-positive subpopulation is not sufficiently more effective than the full population, then the trial continues with the full population. Hence, as per the selection rule in Section 7.3 the condition of selection is now defined by $Q = \hat{\theta}_{1S} \geq \hat{\theta}_{1S^c} - c$. In this case, follow-up in stage 2 continues in both S and S^c with new patients randomised to the experiential treatment and control arms in both subgroups. Hence, an estimate for the subpopulation complement, S^c, is now also observed, where interest at the final analysis is in the overall treatment effect in the full population, θ_F. However, if the full population is selected, θ_S and θ_{Sc} both exist which means the odds are proportional for both S

and S^c and therefore the proportionality assumption cannot hold for F unless the following two conditions hold: i) there is no difference in the hazard rate of S and S^c and ii) the hazard ratio is the same in S and S^c when F is selected.

7.5.1 Naive estimators

A naive estimator for θ_F is derived as the pooled estimate of each subpopulation effect weighted by the prevalence. Conditional on selection, naive estimators at the interim and final analysis in each subpopulation k are $\hat{\theta}_{1k} \sim N(\theta_k, \frac{1}{V_{1k}})$ and $\hat{\theta}_{2k} \sim N(\theta_k, \frac{1}{V_{2k}^F})$. Recall from equation (7.2), $\tilde{\theta}_{2S}^F \sim N(\theta_S, \sigma_{2S^F}^2)$ and $\tilde{\theta}_{2S^c}^F \sim N(\theta_{S^c}, \sigma_{2S^cF}^2)$ are the stage 2 independent increment estimates for subgroup S and S^c, which account for the correlation between stage 1 and stage 2 estimates. As the trial continues with the full population, the information fraction for the subpopulation k is $\tau_k^F = \frac{V_{1k}}{V_{2k}^F}$ $(k = S, S^c)$. Thus, for the biomarker-positive subgroup, a naive estimator is defined by

$$\hat{\theta}_S^F = \tau_S^F \hat{\theta}_{1S} + (1 - \tau_S^F)\tilde{\theta}_{2S}^F$$

and for the biomarker-negative subgroup, a naive estimator is defined by

$$\hat{\theta}_{S^c}^F = \tau_{S^c}^F \hat{\theta}_{1S^c} + (1 - \tau_{S^c}^F)\tilde{\theta}_{2S^c}^F.$$

Hence, the naive estimator when the full population is selected is given by

$$\hat{\theta}_F = p_S \hat{\theta}_S^F + p_{S^c} \hat{\theta}_{S^c}^F. \tag{7.11}$$

Since p_S and p_{S^c} change over time based on the hazard rate, the effect in the full population is calculated at time zero, that is, at the start of trial.

7.5.2 Asymptotically unbiased estimators

When the full population is selected, treatment efficacy is estimated for both the subpopulation, S, and its complement, S^c. Following equation (7.3), conditional on Q, stage 2 increment estimates $\tilde{\theta}^F_{2S}$ and $\tilde{\theta}^F_{2S^c}$ are unbiased for θ_S and θ_{S^c}, respectively. Thus, by the technique of Rao-Blackwellisation of these estimates, derivation of the UMVCUEs for θ_S and θ_{S^c} follow the same steps as described in Section 7.4.2, where sufficient and complete statistics are obtained based on the joint density

$$\begin{pmatrix} \hat{\theta}_{1S} \\ \hat{\theta}_{1S^c} \\ \tilde{\theta}^F_{2S} \\ \tilde{\theta}^F_{2S^c} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_S \\ \theta_{S^c} \\ \theta_S \\ \theta_{S^c} \end{pmatrix}, \begin{pmatrix} \frac{1}{V_{1S}} & 0 & 0 & 0 \\ 0 & \frac{1}{V_{1S^c}} & 0 & 0 \\ 0 & 0 & \frac{1}{V^F_{2S}-V_{1S}} & 0 \\ 0 & 0 & 0 & \frac{1}{V^F_{2S^c}-V_{1S^c}} \end{pmatrix} \right). \tag{7.12}$$

Thus from this density, and following the steps in the previous section, we obtain estimates $\tilde{\theta}^*_{S^F} = \frac{\sigma_{2S^F}}{\sigma_{1S}}\hat{\theta}_{1S} + \frac{\sigma_{1S}}{\sigma_{2S^F}}\tilde{\theta}^F_{2S}$ and $\tilde{\theta}^*_{S^c} = \frac{\sigma_{2S^cF}}{\sigma_{1S^c}}\hat{\theta}_{1S^c} + \frac{\sigma_{1S^c}}{\sigma_{2S^cF}}\tilde{\theta}^F_{2S^c}$, which are sufficient and complete for θ_S and θ_{S^c}. Hence, by Rao-Blackwellisation, the UMVCUE for θ_S when F is selected is

$$\tilde{\theta}^F_S = E[\tilde{\theta}^F_{2S}|\tilde{\theta}^*_{S^F}, \hat{\theta}_{1S^c}, Q] = \hat{\theta}^F_S - \frac{\sigma^2_{2S^F}}{\sqrt{\sigma^2_{1S}+\sigma^2_{2S^F}}}\frac{\phi(\omega^F_S)}{\Phi(\omega^F_S)}, \tag{7.13}$$

where $\omega^F_S = \frac{\sqrt{\sigma^2_{1S}+\sigma^2_{2S^F}}}{\sigma^2_{1S}}\left\{\hat{\theta}^F_S - (\hat{\theta}_{1S^c} - c)\right\}$.

Similarly for θ_{S^c}, conditional on Q, and sufficient and complete statistics $(\hat{\theta}_{1S}, \tilde{\theta}^*_{S^c})$, the UMVCUE for θ_{S^c} is given by

$$\tilde{\theta}^F_{S^c} = E[\tilde{\theta}^F_{2S^c}|\tilde{\theta}^*_{S^c}, \hat{\theta}_{1S}, Q] = \hat{\theta}^F_{S^c} + \frac{\sigma^2_{2S^cF}}{\sqrt{\sigma^2_{1S^c}+\sigma^2_{2S^cF}}}\frac{\phi(\omega_{S^c})}{\Phi(\omega_{S^c})}, \tag{7.14}$$

where $\omega_{S^c} = \frac{\sqrt{\sigma^2_{1S^c}+\sigma^2_{2S^cF}}}{\sigma^2_{1S^c}} \left\{\hat{\theta}_{1S} - c - \hat{\theta}^F_{S^c}\right\}$.

Therefore, an unbiased estimator for θ_F is the weighted sum of the UMVCUEs obtained for each subpopulation, weighted by the subpopulation prevalences at time zero, that is,

$$\tilde{\theta}_F = p_S\tilde{\theta}^F_S + p_{S^c}\tilde{\theta}^F_{S^c}. \tag{7.15}$$

7.6 Simulation study

7.6.1 Simulation set up

The properties of the estimators derived in this chapter are now compared in a simulation study. We envisage a two-stage trial with one experimental treatment compared to a control arm, with randomisation stratified by subpopulation status. We assess the influence of the treatment effect size, the subpopulation prevalence and the selection time on bias and RMSE of the naive and unbiased estimators. The simulation study is set up as described in Section 6.9.1 with survival times simulated from a Weibull distribution corresponding to a range of true log HRs from -0.92 to 0. Since the simulation study in the previous chapter found selection bias to be greatest for equal effect sizes, we assume a common treatment effect in S and S^c, that is, $\theta_S = \theta_{S^c}$. A range of subpopulation prevalences $p_S \in \{0.4, 0.5, 0.7\}$ are considered for selection times $\tau_S \in \{0.3, 0.5, 0.7\}$, which is the information fraction for subpopulation S.

In each simulated study, at the interim analysis estimates $\hat{\theta}_{1k}$ are obtained for each subgroup k ($k = S, S^c$) based on a stratified log HR, as described in equation (4.11). The trial continues with the either the subpopulation S or the full population F. We assume the selection margin $c = \frac{b}{p_S}$ is equal to 0, which is considered the worst

case scenario. Thus, if $\hat{\theta}_{1S} < \hat{\theta}_{1S^c}$ subpopulation S is selected, otherwise F is selected.

The total number of events at the final analysis is set to $D = 1500$ with an interim analysis conducted after d events, where $d = \tau_S D$. Additionally, as in the previous chapter, the number of delayed events is defined as $D_{\text{follow}} = d + 0.1D$. In the case where S is selected, the final analysis is conducted after D events have been observed in subpopulation S. In the case where F is selected, the final analysis is conducted after D events have been observed over both subgroups S and S^c.

At the final analysis, estimates are computed as described by equations (7.5) and (7.10), if S is selected, and equations (7.11) and (7.15), if F is selected.

7.6.2 Simulation results

In the case where S is selected, Table 7.2 presents the average bias and RMSE from 10,000 simulated studies of the naive estimator $\hat{\theta}_S$ and the UMVCUE $\tilde{\theta}_S$. It can be seen that $\tilde{\theta}_S$ is unbiased for small effect sizes, that is, a HR of 0.8 or greater. Bias improves as the subpopulation prevalence increases, however there is a slight increase in bias with selection time τ_S. This is expected as we observed in the previous chapter, bias increases as the interim analysis is performed later in the trial. Additionally, for $\theta_S \geq -0.22$, we find the UMVCUE corrects for the overestimation of the treatment effect in the naive estimator; as can be seen there is a small positive bias which increases with both subpopulation prevalence and selection time.

For large effect sizes, bias in the naive estimator is found to be equivalent to bias in the new estimator; for example, in the case where $p_S = 0.7$, for $\theta_S = -0.92$, $\hat{\theta}_S$ and $\tilde{\theta}_S$ both overestimate the true effect size by 0.0293. Additionally, comparable to the simulation results in Chapter 6 for treatment selection, the RMSE of the

UMVCUE is larger than that of the naive estimator. These results indicate that the asymptotic normality approximation of these estimators are not appropriate for such large effect sizes. This is because the normality approximation is only reasonable under small effect sizes when groups have equal sample sizes. Therefore, as the effect size increases the normality approximation fails due to the tailor series approximation around the log HR of zero. The further we move away from this the worse the approximation is.

If F is selected, Table 7.3 gives the average bias and RMSE from 10,000 simulations of $\hat{\theta}_F$ and $\tilde{\theta}_F$. Similar results are observed to the case when S is selected for large effect sizes. It can be seen the properties of the naive estimator and UMVCUE are comparable. However, the UMVCUE is only asymptotically unbiased under the null, that is, if $\theta_S = 0$. The underestimation observed in both estimators for $\theta_S \neq 0$ is due to the non-constant hazard ratio in the full population. As discussed in Section 7.5, when F is selected, proportional hazards in each subpopulation means the proportionality assumption no longer holds in the full population.

7.7 Conclusion

This chapter aimed to address the issues with estimation in two-stage confirmatory trials with subpopulation selection. The problem of correlated stagewise estimates due to censoring at the interim analysis was addressed by utilising independent increments. To correct for the bias due to selection, by extension of the methods by Kimani et al. [2015], UMVCUEs were derived by Rao-Blackwellisation of the stage 2 increment estimates, conditional on selection. The subpopulation selection rule considered a margin of benefit, where the subpopulation S is only selected if it is sufficiently more effective than the full population by some margin c.

If the experimental treatment is not sufficiently more effective in the biomarker-

Table 7.2: Mean bias and RMSE from 10,000 simulations of the naive estimator ($\hat{\theta}_S$) and the UMVCUE ($\tilde{\theta}_S$) for the case where S is selected. A range of log HRs θ_S are considered for subpopulation prevalence $p_S \in \{0.4, 0.5, 0.7\}$, selection times $\tau_S \in \{0.3, 0.5, 0.7\}$ and selection margin $c = 0$.

			Bias				RMSE			
			θ_S:	0	-0.22	-0.51	-0.92	0	-0.22	-0.51	-0.92
p_S	τ_S	Estimator	HR:	1	0.8	0.6	0.4	1	0.8	0.6	0.4
0.4	0.3	$\hat{\theta}_S$		-0.0077	-0.0073	-0.0164	-0.0322	0.0473	0.0463	0.0519	0.0613
		$\tilde{\theta}_S$		0.0051	-0.0001	-0.0141	-0.0321	0.0511	0.0485	0.0524	0.0613
	0.5	$\hat{\theta}_S$		-0.0123	-0.0029	-0.0201	-0.0445	0.0466	0.0504	0.0530	0.0684
		$\tilde{\theta}_S$		0.0034	0.0080	-0.0187	-0.0445	0.0521	0.0577	0.0536	0.0684
	0.7	$\hat{\theta}_S$		-0.0150	-0.0076	-0.0244	-0.0428	0.0492	0.0538	0.0567	0.0686
		$\tilde{\theta}_S$		0.0079	0.0058	-0.0238	-0.0428	0.0613	0.0670	0.0571	0.0686
0.5	0.3	$\hat{\theta}_S$		-0.0077	-0.0054	-0.0092	-0.0096	0.0464	0.0458	0.0481	0.0504
		$\tilde{\theta}_S$		0.0056	0.0012	-0.0078	-0.0096	0.0508	0.0484	0.0487	0.0504
	0.5	$\hat{\theta}_S$		-0.0095	-0.0024	-0.0199	-0.0333	0.0444	0.0489	0.0507	0.0604
		$\tilde{\theta}_S$		0.0066	0.0062	-0.0194	-0.0333	0.0517	0.0551	0.0511	0.0604
	0.7	$\hat{\theta}_S$		-0.0144	-0.0054	-0.0216	-0.0400	0.0459	0.0506	0.0521	0.0634
		$\tilde{\theta}_S$		0.0063	0.0043	-0.0214	-0.0400	0.0567	0.0616	0.0523	0.0634
0.7	0.3	$\hat{\theta}_S$		-0.0068	-0.0032	0.0023	0.0137	0.0451	0.0440	0.0448	0.0484
		$\tilde{\theta}_S$		0.0058	0.0018	0.0029	0.0137	0.0500	0.0463	0.0453	0.0484
	0.5	$\hat{\theta}_S$		-0.0072	-0.0035	-0.0001	0.0053	0.0426	0.0427	0.0432	0.0447
		$\tilde{\theta}_S$		0.0084	0.0007	0.0000	0.0053	0.0512	0.0455	0.0433	0.0447
	0.7	$\hat{\theta}_S$		-0.0094	0.0005	-0.0176	-0.0293	0.0412	0.0480	0.0466	0.0535
		$\tilde{\theta}_S$		0.0090	0.0063	-0.0175	-0.0293	0.0525	0.0554	0.0466	0.0535

Table 7.3: Mean bias and RMSE from 10,000 simulations of the naive estimator ($\hat{\theta}_F$) and the unbiased estimator ($\tilde{\theta}_F$) for the case where F is selected. A range of log HRs θ_F are considered for subpopulation prevalence $p_S \in \{0.4, 0.5, 0.7\}$, selection times $\tau_S \in \{0.3, 0.5, 0.7\}$ and selection margin $c = 0$.

			Bias				RMSE			
			θ_F:	0	-0.22	-0.51	-0.92	0	-0.22	-0.51	-0.92
p_S	τ_S	Estimator	HR:	1	0.8	0.6	0.4	1	0.8	0.6	0.4
0.4	0.3	$\hat{\theta}_F$		-0.0013	0.1087	0.2508	0.4584	0.0233	0.1110	0.2519	0.4590
		$\tilde{\theta}_F$		0.0017	0.1137	0.2572	0.4650	0.0243	0.1162	0.2584	0.4656
	0.5	$\hat{\theta}_F$		-0.0016	0.1109	0.2524	0.4573	0.0216	0.1134	0.2534	0.4578
		$\tilde{\theta}_F$		0.0019	0.1204	0.2607	0.4644	0.0234	0.1234	0.2619	0.4651
	0.7	$\hat{\theta}_F$		-0.0016	0.1128	0.2542	0.4562	0.0203	0.1152	0.2550	0.4568
		$\tilde{\theta}_F$		0.0024	0.1274	0.2639	0.4633	0.0227	0.1311	0.2650	0.4640
0.5	0.3	$\hat{\theta}_F$		-0.0002	0.1149	0.2631	0.4771	0.0214	0.1169	0.2640	0.4776
		$\tilde{\theta}_F$		-0.0002	0.1184	0.2704	0.4890	0.0214	0.1203	0.2713	0.4895
	0.5	$\hat{\theta}_F$		0.0001	0.1178	0.2643	0.4771	0.0197	0.1199	0.2651	0.4775
		$\tilde{\theta}_F$		0.0000	0.1254	0.2757	0.4952	0.0201	0.1275	0.2766	0.4957
	0.7	$\hat{\theta}_F$		0.0003	0.1194	0.2648	0.4758	0.0189	0.1214	0.2656	0.4763
		$\tilde{\theta}_F$		0.0003	0.1321	0.2801	0.4993	0.0198	0.1343	0.2810	0.4998
0.7	0.3	$\hat{\theta}_F$		0.0027	0.1001	0.2268	0.4106	0.0277	0.1039	0.2284	0.4115
		$\tilde{\theta}_F$		-0.0030	0.0951	0.2231	0.4051	0.0306	0.1000	0.2252	0.4063
	0.5	$\hat{\theta}_F$		0.0029	0.1016	0.2281	0.4102	0.0250	0.1049	0.2295	0.4110
		$\tilde{\theta}_F$		-0.0040	0.0961	0.2231	0.4003	0.0297	0.1009	0.2253	0.4017
	0.7	$\hat{\theta}_F$		0.0030	0.1079	0.2291	0.4091	0.0240	0.1113	0.2304	0.4098
		$\tilde{\theta}_F$		-0.0050	0.0982	0.2227	0.3919	0.0302	0.1068	0.2254	0.3939

positive subpopulation, then the trial continues to the second stage with the full population. As the effect sizes are not equal in both subgroups, we cannot directly estimate the effect in the full population assuming a single parameter model. Therefore, for the case when F is selected, UMVCUEs were derived for each subgroup separately and then the treatment effect in the full population was obtained by the weighted sum of the UMVCUEs for each subpopulation, weighted by the subpopulation prevalences.

Properties of the estimators assessed in a simulation study confirmed the estimators are only asymptotically unbiased for small effect sizes. Furthermore, for large effect sizes, simulation results indicated comparable properties of the naive estimator to the UMVCUE. For the case when S is selected, these results are due to failure of the normality approximation of the log HR for large effect sizes. For the case when F is selected, a non-constant hazard ratio in the full population leads to underestimated effect estimates. This is because the proportionality assumption in the full population no longer holds with proportional hazards in each subpopulation.

Chapter 8

Summary

This thesis focussed on point estimation in two-stage confirmatory clinical trials. Asymptotically conditionally unbiased estimators for time-to-event data were developed for two specific settings; the first focussed on treatment selection in seamless phase II/III clinical trials, and the second focussed on subpopulation selection in the setting of two-stage targeted therapy trials.

In the former setting of treatment selection, multiplicity issues arise when multiple treatments are investigated simultaneously. At an interim analysis, based on a predefined selection criterion, the most efficacious treatment is selected to take forward into the second stage. This selection leads to biased estimation at the final analysis, as stage 1 data are used for both treatment selection as well as for the confirmatory analysis, thus resulting in an overly optimistic estimate. There is therefore a need to correct for this bias in order to obtain an unbiased and thus reliable estimate of the treatment effect.

For normally distributed outcomes, methods have been proposed that correct for this selection bias, as presented in Chapter 5. These utilise the technique of Rao-Blackwellisation, which involves conditioning on the selection rule and complete,

sufficient statistics to derive a UMVCUE. It was shown that the bias and RMSE of a naive maximum likelihood estimator (which ignores the selection rule) depend on several factors including, the effect size, the selection time and the number of experimental treatments under investigation. When compared by simulation to the UMVCUE proposed by Kimani et al. [2013], it was found that, although unbiased, the UMVCUE had a larger RMSE, depending on the selection time. Therefore, as concluded by several authors in the literature, for normally distributed data, a trade-off of bias and RMSE should be considered when utilising UMVCUE's for the analysis of normally distributed data in seamless phase II/III clinical trials.

Although normalised test statistics may be obtained for time-to-event outcomes, existing methods of unbiased estimation cannot be applied directly due to the inherent nature of censoring in time-to-event data. In the setting of seamless phase II/III clinical trials, censored events at the interim analysis lead to correlated stagewise data, which were assumed independent in the setting with normally distributed data. In addition, comparison with a common control arm results in correlated stage 1 estimates. Chapter 6 therefore addressed these issues by developing new unbiased estimators that account for both treatment selection bias and correlation for time-to-event outcomes.

A simulation study investigating the degree of selection bias and RMSE of the naive estimator, which disregards the selection rule, was presented in Section 6.3. This study assessed the influence of the effect size and information fraction, as well as the number treatment arms, and indicated that the naive estimator is positively biased towards a large effect size. This means that the naive estimator overestimates the true treatment effect, where it was shown that bias and RMSE increase with the information fraction (proportion of events observed in stage 1). Furthermore, when the effect sizes of all treatments are equal, it was concluded that, as the number of experimental treatment arms increases, bias and RMSE also increase. However,

when there is a clear superior treatment, bias was found to be roughly constant for any selection time, which was similarly observed for later selection times in the case of normally distributed outcomes (see Figure 5.1(c)). Interestingly, however, for the case of one experimental treatment, where the naive estimator would be expected to be unbiased, it was found to be slightly negatively biased, that is, it underestimates the true effect size, possibly due to the correlation from censored observations at the interim analysis. This simulation study therefore suggested that in addition to the selection rule, the correlation between stage 1 and 2 data should also be accounted for in the derivation of unbiased estimators.

To account for this correlation, the concept of independent increments was presented. This idea, first introduced by Tsiatis [1982], suggests that increment statistics may be defined by the difference between correlated stagewise score statistics, in order to give an independent covariance structure in the joint distribution of stagewise statistics. This is most commonly known as the 'independent increment structure' and has allowed group sequential tests in seamless phase II/III clinical trials to be based on the log-rank test statistic. Therefore, by defining defining increment log HRs for stage 2, we derived the joint distribution of stage 1 and 2 log HRs with independent stagewise test statistics. This solved the problem of the correlation between stage 1 and stage 2 data, however the correlation between stage 1 statistics due to the common control arm remains, as in the case with normal data.

We first considered uncorrelated stage 1 test statistics by assuming a separate control group for each experimental treatment group, which led to the extension of the Kimani et al. [2013] estimator. We incorporated the the independent increment structure and used the method of Rao-Blackwellisation to derive a UMVCUE accounting for selection of the minimum log HR and correlation between stagewise test statistics. Since the stage 2 increment estimate is unbiased, we sought complete,

sufficient statistics in order to Rao-Blackwellise the stage 2 increment estimate by conditioning on the selection rule and the complete, sufficient statistics. This resulted in an estimator of a similar form to the Kimani et al. [2013] and Bowden and Glimm [2008] estimators, but with the difference of allowing for stage 1 and 2 estimates to be correlated by continuing follow-up of patients censored at the interim analysis. Therefore, our estimator corrects for both the bias due to selection of the smallest effect size and the correlation due to censoring. The first component of this estimator, which includes the increment log HR, is equivalent to the naive estimator discussed in the motivational simulation study. Hence, this estimator corrects for the overestimation of the naive estimator.

Since in practice it is common for a trial to include a common control arm, we extended the previous estimator to correct for the correlation between stage 1 estimates, which arises due to the common control arm. Adapting the method by Robertson et al. [2016], a UMVCUE was derived to correct for selection bias as well as both the correlation between stage 1 statistics and the correlation between stage 1 and stage 2 data. As before, this estimator also corrects for the overestimation of the naive estimator but in this case the overcorrection term considers the correlation between the ordered stage 1 log HRs.

As these estimators were derived based on the assumption of asymptotic normality of the log-rank test statistic, we may only conclude that these estimators are asymptotically unbiased. A simulation study was therefore conducted to assess the performance of these estimators for a realistic sample size with simulated survival times. A comparison of the naive estimator with the new UMVCUE highlighted favourable properties of the UMVCUE under certain scenarios. Specifically, due to normality approximation of the log-rank test statistic, simulation results showed that this estimator is only asymptotically unbiased for small to moderate effect sizes, but for any number of experimental treatment arms and selection times.

Chapter 7 presented the issues in estimation following subpopulation selection in targeted therapy trials. The need for unbiased and efficient estimators in this setting led to the development of UMVCUEs for time-to-event data correcting for subgroup selection and correlation. As in the setting of treatment selection, the problem of correlated test statistics arises when, at the end of stage 1, patients who do not experience the event of interest are censored and followed-up further in the second stage. Therefore, by extension of the methods by Kimani et al. [2015], a UMVCUE was derived by Rao-Blackwellisation for each possible case of subgroup selection. If the experimental treatment is not sufficiently more effective in the biomarker-positive subpopulation, the trial continues to the second stage with the full population. In this case, UMVCUEs were obtained for each subgroup separately and the effect in the full population was estimated by the weighted sum of the UMVCUEs for each subgroup. Properties of the estimators assessed in a simulation study showed favourable properties of the UMVCUE in the case where S is selected, for small effect sizes. However, in the case where F is selected, simulation results indicated comparable properties of the UMVCUE to the naive estimator.

In conclusion, this thesis developed asymptotically uniformly minimum variance unbiased estimators, conditional on selection rules, for use in the analysis of time-to-event data. Due to the asymptotic assumptions, these estimators are only appropriate for large samples with small to moderate effect sizes. As phase III trials are confirmatory in nature, such that they are designed with large numbers of events in order to achieve adequate power to detect large effect sizes, the asymptotic assumptions of these estimators would be satisfied under these conditions. These methods therefore contribute to the statistical methodology for unbiased and efficient estimation of time-to-event outcomes in the setting of two-stage confirmatory clinical trials with treatment selection or subpopulation selection.

8.1 Further work

The following list suggests areas for further work.

(i) In Chapter 7, for the case where the full population is selected to continue in the second stage, an unbiased estimator was presented based on the weighted sum of the UMVCUEs for each of the biomarker-positive and biomarker-negative subpopulations. It would be of interest to directly estimate the treatment effect from the full population, rather than weighting estimators by the subpopulation prevalence.

The asymptotic joint distribution under the null of stage 1 and 2 estimates for estimating the effect in the full population directly is

$$\begin{pmatrix} \hat{\theta}_{1S} \\ \hat{\theta}_{1F} \\ \tilde{\theta}_{2F} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_S \\ \theta_F \\ \theta_F \end{pmatrix}, \begin{pmatrix} \frac{1}{V_{1S}} & \frac{p_S}{V_{1S}} & 0 \\ \frac{p_S}{V_{1S}} & \frac{1}{V_{1F}} & 0 \\ 0 & 0 & \frac{1}{V_{2F}-V_{1F}} \end{pmatrix} \right),$$

where the correlation between stage 1 statistics is $p_S\sqrt{\frac{v_{1F}}{v_{1S}}}$.

However, this density only holds if the effects in each subpopulation are equal, that is, if $\theta_S = \theta_{S^c}$. This is because it is not appropriate to fit a model with a single parameter when the effects in each subgroup are expected to be different. Therefore, further work is needed to derive the joint density of the estimators for the effect in the biomarker-positive subpopulation in stage 1 and the full-population estimate from both stages. It would then be of interest to compare the results from the weighted estimator given in equation (7.15) with the estimator obtained from the correct joint density.

(ii) Brückner et al. [2017] recently proposed a shrinkage estimator based on the asymptotic normality approximation of the log HR obtained from a Cox PH

model for time-to-event data. Their estimator accounts for treatment selection in two-stage trials. Brückner et al. compare the naive estimator obtained at the final analysis with both their shrinkage estimator and the estimator proposed by Stallard and Todd [2005] for normally distributed data. Their estimator showed favourable properties when compared to the Stallard and Todd estimator, in particular for MSE. In a similar way to Brückner et al. it would worthwhile to develop a shrinkage estimator, but based on the log-rank test statistic. This estimator can then be compared to the UMVCUE derived in Section 6.8.

(iii) As the main assumption of the estimators developed in this thesis is that of asymptotic normality of the log HR, further work to develop estimators specific to survival distributions would be of value. Shen and Cheng [2006] provide an estimator for the simple case where one experimental treatment is compared to a control and the only adaptation at the interim analysis is that of early stopping for futility or efficacy. This method may be extended to include a predefined selection rule, either for the case of treatment selection, where two or more experimental treatments are investigated in the first stage with the aim of selecting the best performing treatment, or for the setting of subpopulation selection. A uniformly minimum variance unbiased estimator, conditional on selection, for specific survival distributions could then be derived. These estimators could then be compared to those developed in this thesis for time-to-event data.

(iv) Results from simulated data were presented in this thesis in order to compare the estimators developed under various trial scenarios. It would be of value to apply these methods to real trial data in order to confirm the applicability of these estimators for a real trial setting, where problems may arise such as failure of the proportional hazards assumption.

(v) In general, novel statistical methods are not frequently used in practice due to the lack of readily available software. Therefore, it would be beneficial to develop an R package to aid implementation of these estimators. An example of this is the 'asd' package [Parsons, 2016], which provides functions for hypothesis testing in adaptive seamless phase II/III clinical trials with treatment selection.

Appendix A

Proof of expectation of a truncated normal

Let X have a normal distribution with mean μ and variance σ^2, then the expectation of a truncated normal is

$$\frac{1}{\sigma}\int_{-\infty}^{b} x\phi\left(\frac{x-\mu}{\sigma}\right)\mathrm{d}x = -\sigma\phi\left(\frac{b-\mu}{\sigma}\right) + \mu\Phi\left(\frac{b-\mu}{\sigma}\right).$$

Proof.

Consider the moment generating function of X

$$m(t) = \frac{1}{\sigma}\int_{b}^{-\infty} e^{tx}\phi\left(\frac{x-\mu}{\sigma}\right)\mathrm{d}x$$

The first partial derivative of $m(t)$ is given by

$$\begin{aligned}\frac{\partial m}{\partial t} &= \frac{1}{\sigma}\int_{-\infty}^{b}\frac{\partial}{\partial t}e^{tx}\phi\left(\frac{x-\mu}{\sigma}\right)\mathrm{d}x \\ &= \frac{1}{\sigma}\int_{-\infty}^{b} xe^{tx}\phi\left(\frac{x-\mu}{\sigma}\right)\mathrm{d}x\end{aligned}$$

so that when $t = 0$,

$$\frac{\partial m}{\partial t} = \frac{1}{\sigma}\int_{-\infty}^{b} x\phi\left(\frac{x-\mu}{\sigma}\right)\mathrm{d}x.$$

Now,

$$\begin{aligned} m(t) &= \frac{1}{\sigma}\int_{-\infty}^{b} e^{tx}\phi\left(\frac{x-\mu}{\sigma}\right)\mathrm{d}x \\ &= \int_{-\infty}^{b}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}+tx\right\}\mathrm{d}x. \end{aligned}$$

Expanding terms in the exponent and completing the square, we can write

$$\begin{aligned} & -\frac{(x-\mu)^2}{2\sigma^2}+tx \\ = \; & -\frac{1}{2\sigma^2}(x^2-2\mu x+\mu^2)+tx \\ = \; & -\frac{1}{2\sigma^2}(x^2-2x(\mu+\sigma^2t)+\mu^2) \\ = \; & -\frac{1}{2\sigma^2}(x^2-2x(\mu+\sigma^2t)+(\mu+\sigma^2t)^2-(\mu+\sigma^2t)^2+\mu^2) \\ = \; & -\frac{1}{2\sigma^2}(x^2-2x(\mu+\sigma^2t)+(\mu+\sigma^2t)^2)+\frac{1}{2\sigma^2}(\mu^2+2\mu\sigma^2t+\sigma^4t^2-\mu^2) \\ = \; & -\frac{1}{2\sigma^2}(x-(\mu+\sigma^2t))^2+\mu t+\frac{\sigma^2t^2}{2} \end{aligned}$$

and so,

$$\begin{aligned} m(t) &= \int_{-\infty}^{b}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}+tx\right\}\mathrm{d}x \\ &= \exp\left\{\mu t+\frac{\sigma^2t^2}{2}\right\}\int_{-\infty}^{b}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-(\mu+\sigma^2t))^2\right\}\mathrm{d}x \\ &= \exp\left\{\mu t+\frac{\sigma^2t^2}{2}\right\}\Phi\left(\frac{b-(\mu+\sigma^2t)}{\sigma}\right). \end{aligned}$$

Hence,

$$\frac{\partial m}{\partial t} = (\mu + \sigma^2 t) \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \Phi\left(\frac{b - (\mu + \sigma^2 t)}{\sigma}\right) \\ -\sigma \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \phi\left(\frac{b - (\mu + \sigma^2 t)}{\sigma}\right)$$

and setting $t = 0$ gives,

$$\frac{\partial m}{\partial t} = \mu \Phi\left(\frac{b - \mu}{\sigma}\right) - \sigma \phi\left(\frac{b - \mu}{\sigma}\right).$$

□

Appendix B

Proof of sufficiency and completeness of $\tilde{\boldsymbol{\theta}}^*$

This proof follows from Robertson et al. [2016]. For simplicity, the notation from Section 6.8 is changed as follows.

Let the vector of stage 1 log HR's, $\hat{\boldsymbol{\theta}}$, be denoted by $\boldsymbol{X} = (X_1, ..., X_K)$ with unknown true log HRs $\boldsymbol{\mu} = (\mu_1, ..., \mu_K)$ and covariance $\boldsymbol{V} = V_{ij}$ for $i, j \in \{1, ..., K\}$. Let Y denote the stage 2 increment statistic $\tilde{\theta}_{21}$ for the selected treatment such that $Y \sim N(\mu, \tau^2)$. Finally, let $\boldsymbol{Z} = (Z_1, ..., Z_K)$ denote the sufficient and complete statistics $\tilde{\boldsymbol{\theta}}^*$, which we want to show are sufficient and complete for $\boldsymbol{\theta}$.

Proof.

Recall the selection rule Q_s defined in Section 6.2. Following Equation (6.20) the joint distribution of $(\boldsymbol{X}, Y)$ given Q_s has the density

$$f(\boldsymbol{x}, y|Q_s) = \frac{1}{P_{Q_s}} I_{Q_s}(\boldsymbol{x}) g(\boldsymbol{x}) \frac{1}{\tau} \phi\left(\frac{y - \mu_1}{\tau}\right) \tag{B.1}$$

where, $I_{Q_s}(\boldsymbol{x})$ is the indicator function for Q_s, $P(\boldsymbol{x}) = Prob(I_{Q_s}(\boldsymbol{x}) = 1)$ and

$$g(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^K |\boldsymbol{V}|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})' P (\boldsymbol{x}-\boldsymbol{\mu})\right\} \tag{B.2}$$

Now the exponent of $f_{Q_s}(x, y)$ is

$$\begin{aligned} & \frac{1}{2}\left[\boldsymbol{x}'P\boldsymbol{\mu} + \boldsymbol{\mu}'P\boldsymbol{x} - \boldsymbol{\mu}'P\boldsymbol{\mu} - \boldsymbol{x}'P\boldsymbol{x} - \frac{1}{\tau^2}y^2 + \frac{2}{\tau^2}y\mu_1 - \frac{1}{\tau^2}\mu_1^2\right] \\ = & \frac{1}{2}\left[2\left(\sum_{i=1}^K p_{1i}x_i + \frac{1}{\tau^2}y\right)\mu_1 + 2\sum_{i=2}^K\left(\sum_{j=1}^K p_{ij}x_j\right)\mu_i - \boldsymbol{\mu}'P\boldsymbol{\mu} - \frac{1}{\tau^2}\mu_1^2 + \psi(\boldsymbol{x}, y)\right] \end{aligned} \tag{B.3}$$

where $\psi(\boldsymbol{x}, y) = -\boldsymbol{x}'P\boldsymbol{x} - \frac{1}{\tau^2}y^2$.

Substituting (B.3) into (B.1) and using the Factorisation criterion (3.15)

$$\begin{aligned} f_{Q_s}(\boldsymbol{x}, y) &= \frac{1}{P_{Q_s}} I_{Q_s}(x) \frac{1}{\sqrt{2\pi|\boldsymbol{V}|}} \exp\left\{\frac{1}{2}\left(\sum_{i=1}^K p_{1i}x_i + \frac{1}{\tau^2}y\right)\mu_1\right. \\ &\qquad \left. + 2\sum_{i=2}^K\left(\sum_{j=1}^K p_{ij}x_j\right)\mu_i + \boldsymbol{\mu}'P\boldsymbol{\mu} - \frac{1}{\tau^2}\mu_1^2 + \psi(\boldsymbol{x}, y)\right\} \\ &= \frac{1}{P_{Q_s}} I_{Q_s}(x) \frac{1}{\sqrt{2\pi|\boldsymbol{V}|}} \exp\left\{\frac{1}{2}\left(\sum_{i=1}^K p_{1i}x_i + \frac{1}{\tau^2}y\right)\mu_1\right. \\ &\qquad \left. + 2\sum_{i=1}^K\left(\sum_{j=1}^K p_{ij}x_j\right)\mu_i\right\} \exp\left\{\boldsymbol{\mu}'\boldsymbol{\mu} - \frac{1}{\tau^2}\mu_1^2 + \psi(\boldsymbol{x}, y)\right\}. \end{aligned} \tag{B.4}$$

Hence, $\boldsymbol{T} = (T_1, \ldots, T_K)$ is sufficient for $\boldsymbol{\mu}$, where

$$\begin{aligned} T_1 &= \sum_{i=1}^K p_{1i}x_i + \frac{1}{\tau^2}y \\ T_i &= \sum_{j=1}^K p_{ij}x_j \qquad\qquad \text{for } i = 1, 2, \ldots, K. \end{aligned}$$

By Definition 3.4.8, it can be concluded that $\boldsymbol{T}$ is also a complete statistic since density (B.4) is in natural parametrisation of an exponential family (3.7) and the natural parameter space contains a $K-$dimensional rectangle.

Now, in order to get the required sufficient statistic, linear combinations of $\boldsymbol{T}$ are formed. First consider if $V_{1i} \neq 0$ for $i = 1, 2, \ldots, K$.

Let

$$\begin{aligned}
\tilde{Z}_i &= T_1 + \sum_{j=2}^{K} \frac{V_{ij}}{V_{1i}} T_j \\
&= \sum_{j=1}^{K} p_{ij} x_j + \frac{1}{\tau^2} y + \frac{1}{V_{1i}} \sum_{j=2}^{K} V_{1j} \sum_{k=1}^{K} p_{jk} x_k \\
&= \sum_{j=1}^{K} p_{ij} x_j + \frac{1}{\tau^2} y + \frac{1}{V_{1i}} \left[\sum_{j,k=1}^{K} V_{1j} p_{jk} x_k - V_{1i} \sum_{k=1}^{K} p_{1k} x_k \right] \\
&= \frac{1}{\tau^2} y + \frac{1}{V_{1i}} \left[V_{1i} \sum_{j=1}^{K} p_{1j} x_j + \sum_{j,k=1}^{K} V_{1j} p_{jk} x_k - V_{li} \sum_{k=1}^{K} p_{1k} x_k \right] \\
&= \frac{1}{\tau^2} y + \frac{1}{V_{1i}} \left[\sum_{j,k=1}^{K} V_{1j} p_{jk} x_k \right] \\
&= \frac{1}{\tau^2} y + \frac{x_i}{V_{1i}} \\
&= \frac{1}{V_{1i}} \left[x_i + \frac{V_{1i}}{\tau^2} y \right]
\end{aligned}$$

$\implies Z_i = x_i + \dfrac{V_{1i}}{\tau^2} Y$ is sufficient for μ_i.

If $V_{1i} = 0$,

$$\begin{aligned}
\tilde{Z}_i &= \sum_{j=2}^{K} V_{1j} T_j \\
&= \sum_{j=2}^{K} V_{1j} \sum_{k=1}^{K} p_{jk} x_k \\
&= \sum_{j,k=1}^{K} V_{1j} p_{jk} x_k \\
&= x_i
\end{aligned}$$

Hence, $Z_i = X_i + \frac{V_{1i}}{\tau^2} Y$ is sufficient for μ_i. Therefore, since $\boldsymbol{Z}$ is a linear transformation of $\boldsymbol{T}$, and $\boldsymbol{T}$ is a complete sufficient statistic then $\boldsymbol{Z}$ is also sufficient and complete.

□

Bibliography

A'Hern, R. P. (2001). Sample size tables for exact single-stage phase II designs. *Statistics in Medicine*, 20:859–866.

Bauer, P., Koenig, F., Brannath, W., and Posch, M. (2010). Selection and bias - two hostile brothers. *Statistics in Medicine*, 29:1 – 13.

Bauer, P. and Posch, M. (2004). Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections by h. schäfer and h.-h. müller, statistics in medicine 2001; 20: 3741–3751. *Statistics in Medicine*, 23(8):1333–1334.

Bowden, J. and Glimm, E. (2008). Unbiased estimation of selected treatment means in two-stage trials. *Biometrical Journal*, 50(4):515–527.

Brannath, W., Zuber, E., Branson, M., Bretz, F., Gallo, P., Posch, M., and Racine-Poon, A. (2009). Confirmatory adaptive designs with bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine*, 28:1445–1463.

Bretz, F., König, F., Brannath, W., Glimm, E., and Posch, M. (2009). Adaptive designs for confirmatory clinical trials. *Statistics in Medicine*, 28:1181–1217.

Bretz, F., König, F., Racine, A., and Maurer, W. (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: General concepts. *Biometrical Journal*, 48:623–634.

Brückner, M., Titman, A., and Jaki, T. (2017). Estimation in multi-arm two-stage trials with treatment selection and time-to-event endpoint. *Statistics in Medicine*, 36(20):3137–3153.

Bryant, J. and Day, R. (1995). Incorporating toxicity considerations into the design of two-stage phase II clinical trials. *Biometrics*, 51:1372–1383.

Buyse, M., Michiels, S., Sargent, D. J., Grothey, A., Matheson, A., and de Gramont, A. (2011). Integrating biomarkers in clinical trials. *Expert Review of Molecular Diagnostics*, 11(2):171–182.

Casella, G. and Berger, R. (2002). *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning.

Chuang-Stein, C., Anderson, K., Gallo, P., and Collins, S. (2006). Sample size re-estimation: A review and recommendations. *Drug Information Journal*, 40:475–484.

Cohen, A. and Sackrowitz, H. B. (1989). Two stage conditionally unbiased estimators of the selected mean. *Statistics and Probability Letters*, 8:273 – 278.

Collett, D. (2015). *Modelling Survival Data in Medical Research, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press.

Conaway, M. R. and Wages, N. A. (2016). *Phase I Trials and Dose-finding*, chapter 6, pages 167–202. In George et al. [2016].

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B*, 34(2):187 – 220.

Desseaux, K. and Porcher, R. (2007). Flexible two-stage design with sample size reassessment for survival trials. *Statistics in Medicine*, 26(27):5002–5013.

Di Scala, L. and Glimm, E. (2011). Time-to-event analysis with treatment arm selection at interim. *Statistics in Medicine*, 30:3067–3081.

Fleming, T. R. (1982). One-sample multiple testing procedure for phase II clinical trials. *Biometrics*, 38:143–151.

Friede, T., Parsons, N., and Stallard, N. (2012). A conditional error function approach for subgroup selection in adaptive clinical trials. *Statistics in Medicine*, 31(30):4309–4320.

George, S. L., Wang, X., and Pang, H. (2016). *Cancer Clinical Trials: Current and Controversial Issues in Design and Analysis*. Chapman & Hall/CRC Biostatistics Series. CRC Press.

Glimm, E. and Di Scala, L. (2015). An approach to confirmatory testing of subpopulations in clinical trials. *Biometrical Journal*, 57(5):897 – 913.

Green, S., Benedetti, J., and Crowley, J. (2002). *Clinical Trials in Oncology*. Chapman & Hall/CRC, second edition.

Hatfield, I., Allison, A., Flight, L., Julious, S. A., and Dimairo, M. (2016). Adaptive designs undertaken in clinical research: a review of registered clinical trials. *Trials*, 17(1):150.

Hommel, G. (2001). Adaptive modifications of hypotheses after an interim analysis. *Biometrical Journal*, 43:581–589.

Irle, S. and Schäfer, H. (2012). Interim design modifications in time-to-event studies. *Journal of the American Statistical Association*, 107(497):341–348.

Jahn-Eimermacher, A. and Ingel, K. (2009). Adaptive trial design: A general methodology for censored time to event data. *Contemporary Clinical Trials*, 30:171 – 177.

Jenkins, M., Stone, A., and Jennison, C. (2011). An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharmaceutical Statistics*, 10:347 – 356.

Jennison, C. and Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC.

Jung, S. H. (2008). Randomized phase II trials with a prospective control. *Statistics in Medicine*, 27:568–583.

Jung, S. H. (2016). *Design and Analysis of Phase II Cancer Trials*, chapter 7, pages 203–231. In George et al. [2016].

Jung, S. H. and Kim, K. M. (2004). On the estimation of the binomial probability in multistage clinical trials. *Statistics in Medicine*, 23:881–896.

Kimani, P. K., Todd, S., and Stallard, N. (2013). Conditionally unbiased estimation in phase II/III clinical trials with early stopping for futility. *Statistics in Medicine*, 32(17):2893–2910.

Kimani, P. K., Todd, S., and Stallard, N. (2015). Estimation after subpopulation selection in adaptive seamless trials. *Statistics in Medicine*, 34:2581 – 2601.

Kleinbaum, D. G. and Klein, M. (2012). *Survival analysis: a self-learning text*. Springer, New York, third edition.

Korn, E. L. and Freidlin, B. (2017). Adaptive clinical trials: Advantages and disadvantages of various adaptive design elements. *JNCI: Journal of the National Cancer Institute*, 109(6):djx013.

Korn, E. L., Freidlin, B., Abrams, J. S., and Halabi, S. (2012). Design issues in randomized phase II/III trials. *Journal of Clinical Oncology*, 30:667–671.

Lakatos, E. (2016). *Sample Size for Survival Trials in Cancer*, chapter 8, pages 235–279. In George et al. [2016].

Lehmann, E. L. and Scheffé, H. (1950). Completeness, similar regions, and unbiased estimation: Part i. *Sankhya: The Indian Journal of Statistics (1933-1960)*, 10(4):305–340.

Lindgren, B. (1993). *Statistical Theory, Fourth Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

Machin, D., Campbell, M. J., Tan, S. B., and Tan, S. H. (2009). *Sample Size Tables for Clinical Tials*. John Wiley & Sons Ltd, Chichester, third edition.

Machin, D., Cheung, Y. B., and Parmar, M. K. (2006). *Survival Analysis: A Practical Approach*. John Wiley & Sons Ltd, Chichester, second edition.

Machin, D., Day, S., and Green, S. (2004). *Textbook of clinical trials*. John Wiley & Sons Ltd, Chichester.

Mehta, C., Schäfer, H., Daniel, H., and Irle, S. (2014). Biomarker-driven population enrichment for adaptive oncology trials with time to event endpoints. *Statistics in Medicine*, 33:4515–4531.

Ondra, T., Dmitrienko, A., Friede, T., Graf, A., Miller, F., Stallard, N., and Posch, M. (2016). Methods for identification and confirmation of targeted subgroups in clinical trials: A systematic review. *Journal of Biopharmaceutical Statistics*, 26(1):99 – 119.

O'Quigley, J., Pepe, M., and Fisher, L. (1990). Continual reassessment method: A practical design for phase 1 clinical trials in cancer. *Biometrics*, 46:33–48.

Parsons, N. (2016). *asd: Simulations for Adaptive Seamless Designs*. R package version 2.2.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society, Series A*, 135:185–207.

Piantadosi, S. (2005). *Clinical Trials: A Methodological Perspective*. John Wiley & Sons Ltd, New Jersey, 2 edition.

Posch, M., Koenig, F., Branson, M., Brannath, W., Dunger-Baldauf, C., and Bauer, P. (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine*, 24:3697–3714.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robertson, D. S., Prevost, A. T., and Bowden, J. (2016). Accounting for selection and correlation in the analysis of two-stage genome-wide association studies. *Biostatistics*, 17(4):634–649.

Roussas, G. (2007). *Introduction to Probability*. Elsevier Academic Press.

Royston, P. and Lambert, P. C. (2011). *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. StataCorp LP.

Rubinstein, L. V., Korn, E. L., Freidlin, B., Hunsberger, S., Ivy, S. P., and Smith, M. A. (2005). Design issues of randomized phase II trials and a proposal for phase ii screening trials. *Journal of Clinical Oncology*, 23:7199–7206.

Sargent, D. J. and Goldberg, R. M. (2001). A flexible design for multiple armed screening trials. *Statistics in Medicine*, 20:1051–1060.

Schäfer, H. and Müller, H.-H. (2001). Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. *Statistics in Medicine*, 20(24):3741–3751.

Sheiner, L. B. (1997). Learning versus confirming in clinical drug development. *Clinical Pharmacology & Therapeutics*, 61(3):275–291.

Shen, L. (2001). Estimation following selection of the largest of two normal means. *Statistics in Medicine*, 20:1913 – 1929.

Shen, Y. and Cheng, Y. (2006). Adaptive design: estimation and inference with censored data in a semiparametric model. *Biostatistics*, 8:306 – 322.

Simon, R. (1989). Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials*, 10:1–10.

Simon, R., Wittes, R. E., and Ellenberg, S. S. (1985). Randomised phase II clinical trials. *Cancer treatment reports*, 69:1375–1381.

Stallard, N. and Todd, S. (2003). Sequential designs for phase III clinical trials incorporating treatment selection. *Statistics in Medicine*, 22:689–703.

Stallard, N. and Todd, S. (2005). Point estimates and confidence regions for sequential trials involving selection. *Journal of Statistical Planning and Inference*, 135:402 – 419.

Stasinopoulos, M., Rigby, B., Akantziliotou, C., and Ospina, R. (2006). *gamlss.dist: Distributions for Generalized Additive Models for Location Scale and Shape*. R package version 1.5.0.

Tarone, R. E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika*, 64(1):156–160.

Tsiatis, A. A. (1982). Repeated significance testing for a general class of statistics used in censored survival analysis. *Journal of the American Statistical Association*, 77(380):855–861.

Wassmer, G. (2006). Planning and analyzing adaptive group sequential survival trials. *Biometrical Journal*, 48(4):714–729.

Wassmer, G. and Brannath, W. (2016). *Group Sequential and Confirmatory Adaptive Designs in Clinical Trials*. Springer Series in Pharmaceutical Statistics. Springer International Publishing.

Whitehead, J. (1997). *The Design and Analysis of Sequential Clinical Trials*. John Wiley & Sons Ltd, England, 2 edition.

Xi, D., Glimm, E., and Bretz, F. (2016). *Multiplicity*, chapter 3, pages 69–104. In George et al. [2016].

Young, G. A. and Smith, R. L. (2005). *Essentials of Statistical Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.